THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

PAGE

Further Contributions to the Problem of Serial Correlation. WILFRID J. DIXON 119

On a Statistical Poblem Arising in the Classification of an Individual into One of Two Groups. ABRAHAM WALD 145

Asymptotic Distribution of Runs Up and Down. J. WOLFOWITZ .. 163

Statistical Analysis of Certain Types of Random Functions. H. HURWITZ AND M. KAC 173

Random Alms. PAUL R. HALMOS 182

On Biases in Estimation Due to the Use of Preliminary Tests of Significance. T. A. BANCROFT 190

The Probability of Convergence of an Iterative Process of Inverting a Matrix. JOSEPH ULLMAN 205

Notes:

On Distribution-free Tolerance Limits in Random Sampling. HERBERT ROBBINS 214

A Formula for Sample Sizes for Population Tolerance Limits. H. SCHEFFE and J. W. TUKEY 217

A Generalization of Waring's Formula. T. N. E. GREVILLE 218

Note on the Variance and Best Estimates. H. G. LANDAU 219

News and Notices 222

Report on the Washington Meeting of the Institute 226

Vol. XV, No. 2 — June, 1944

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves., Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December. Because of war-time difficulties of publication, issues may often be from two to four weeks late in appearing. *Subscribers are therefore requested to wait at least 30 days after month of issue before making inquiries concerning non-delivery.*

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $5.00 per year. Single copies $1.50. Back numbers are available at $5.00 per volume, or $1.50 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the Act of March 3, 1879

FURTHER CONTRIBUTIONS TO THE PROBLEM OF SERIAL CORRELATION

By Wilfrid J. Dixon

Princeton University

1. Introduction. Recently, there has been an increasing interest in the study of the serial correlation of observations. The development of the distribution theory and significance criteria was retarded by the fact that the successive differences or successive products of statistical variates are not independent. However, these difficulties have been overcome to a considerable extent by recent work of several authors. In order to indicate the nature of the contributions embodied in the present paper, it will be necessary to describe rather precisely the contributions of these authors.

Suppose $x_1, x_2, \cdots, x_n$ are n independent observations of a random variable x which is normally distributed with mean a and variance σ^2. Let us define

$$\begin{aligned} \delta^2_{n-1} &= \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 & \delta^2_n &= \sum_{i=1}^{n} (x_{i+1} - x_i)^2 \\ C_{n-1} &= \sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x}) & C_n &= \sum_{i=1}^{n} (x_i - \bar{x})(x_{i+1} - \bar{x}) \\ {}_LC_n &= \sum_{i=1}^{n} (x_i - \bar{x})(x_{i+L} - \bar{x}) & V_n &= \sum_{i=1}^{n} (x_i - \bar{x})^2 \end{aligned} \tag{1.1}$$

in which $x_{n+i} = x_i$. The ratio of any of the first five values to V_n will be a measure of the relation between the successive observations x_i.

Von Neumann [2] has studied the ratio $\eta = \delta^2_{n-1}/V_n$. He obtains an expression for the sampling distribution of the ratio η. He solves the equivalent problem of determining the distribution of $\sum_{i=1}^{n-1} A_i y_i^2$ where the point $(y_1, y_2, \cdots, y_{n-1})$ is uniformly distributed over the spherical surface $\sum_{i=1}^{n-1} y_i^2 = 1$ and the A_i are the characteristic values of δ^2_{n-1}. He obtains the distribution $\omega(\gamma)$ of $\gamma = \sum_{i=1}^{m} B_i x_i^2$ (m even) where the point $(x_1, x_2, \cdots, x_n)$ is uniformly distributed over the spherical surface $\sum_{i=1}^{m} x_i^2 = 1$ and $B_1 \geq B_2 \geq \cdots \geq B_m$. $\omega(\gamma)$ is found by solving the equation

$$\int_{B_m}^{B_1} (\gamma - z)^{-\frac{1}{2}m} \omega(\gamma)\, d\gamma = \prod_{i=1}^{m} (B_i - z)^{-\frac{1}{2}}. \tag{1.2}$$

The distribution of η is then a special case of this distribution. The first four moments were obtained by Williams [5] by the use of a generating function. In

the present paper we shall study the ratio δ_n^2/V_n. The moments of this ratio will be developed and the moments and approximate distribution of $[2 - \delta_n^2/V_n]^2$.

Von Neumann [4] in a paper which removed a restriction (that m be even) on the distribution of η indicates how to determine the distribution of C_{n-1}/V_n. Koopmans [9] considers the stochastical process $x_t = \rho x_{t-1} + z_t$ $(t = 1, 2, \cdots)$, $|\rho| < 1$. The z_t are independent drawings from a normal distribution with zero mean and variance σ^2. To test the hypothesis that $\rho = 0$ he shows that it is sufficient to know the distribution of C_{n-1}/V_n. He finds the distribution of C_{n-1}/V_n and C_n/V_n but finds that the numerical computation of these functions is very cumbersome. This prompts him to obtain approximate formulas for these distributions. The approximate formula for the distribution of $\bar{r} = C_n/V_n$ is

$$(1.3)\qquad (\tfrac{1}{2}n - 1)2^{\frac{1}{2}n}\pi^{-1}\int_0^{\arccos \bar{r}} (\cos\alpha - \bar{r})^{\frac{1}{2}n-2} \sin \tfrac{1}{2}n\alpha \sin\alpha \, d\alpha.$$

A similar approximation will be used in this paper to find the moments of C_n/V_n. It will be shown how good the approximation is and how by using this approximation we may obtain a tabled function (Pearson Type I) which fits the distribution of $1 - (C_n/V_n)^2$ up to $\frac{1}{2}n$ moments.

The quantity $1 - (C_n/V_n)^2$, we shall find, is equivalent to a likelihood ratio function for testing the hypothesis that the serial correlation is zero.

Anderson [8] obtained the distribution of ${}_LC_n/V_n = {}_LR_n$. He proved that the distribution of ${}_LR_n$ is the same as that of ${}_1R_n$ when L and n are prime to each other. He has computed the 1 per cent and 5 per cent significance values $(L = 1)$ up to $n = 75$. For values of $n > 75$ he indicates that a normal distribution which is an asymptotic approximation may be used. He has also computed some significance values for the cases of $N/L = 2, 3, 4$.

In this paper we shall develop the moments of ${}_LR_n$.

The use of the ratio η in the study of serial effects in ballistics at Aberdeen Proving Ground is given in references [1] and [2]. The use of the ratios C_{n-1}/V_n and C_n/V_n in the study of economic time series is discussed by Koopmans [9].

2. Likelihood criteria. Given a sample of n observations $x_1, x_2, \cdots, x_n$ we shall assume that they are distributed according to the law:

$$(2.1)\qquad dP_n = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{\alpha=1}^{n}(x_\alpha - a - bx_{\alpha-l})^2} dx_1 \cdots dx_n, \qquad (1 \le l \le n).$$

It will be convenient to use the phraseology that "the variate at the time α has as its mean value a linear function of the variate at the time $\alpha - l$." We shall take $x_i = x_{n+i}$. Due to the symmetry we may use $\alpha + l$ in place of $\alpha - l$. This will be done to obtain agreement with previous work. We wish to test the hypothesis H_1 that each variate is independent of the other variates, that is, that $b = 0$. The Neyman-Pearson specification of H_1 may be written as follows,

where Ω is the space of admissible values of σ^2, a and b, and ω the subspace defining H_1:

$$
(2.2)\qquad \begin{cases} \Omega: \sigma^2 > 0 & -\infty < a, \quad b < \infty \\ \omega: \sigma^2 > 0 & -\infty < a < \infty, \quad b = 0. \end{cases}
$$

The likelihood criterion λ_1 suitable to this hypothesis is the ratio of the maximum (ω (max.)) of (2.1) with the restriction that $b = 0$ to the maximum (Ω (max.)) of (2.1) without this condition. Now,

$$
(2.3)\qquad \lambda_1 = \frac{dP_n(\omega \max)}{dP_n(\Omega \max)}.
$$

We see that the likelihood function is

$$
(2.4)\qquad L = -n \log (\sqrt{2\pi}\,\sigma) - \frac{1}{2\sigma^2} \sum_{\alpha=1}^{n} (x_\alpha - a - bx_{\alpha+l})^2
$$

and to maximize L over the space Ω we compute the following derivatives

$$
(2.5)\qquad \begin{aligned} \frac{\partial L}{\partial a} &= \frac{1}{\sigma^2} \Sigma(x_\alpha - a - bx_{\alpha+l}), \\ \frac{\partial L}{\partial b} &= \frac{1}{\sigma^2} \Sigma(x_\alpha - a - bx_{\alpha+l})x_{\alpha+l}, \\ \frac{\partial L}{\partial \sigma} &= -\frac{n}{\sigma^2} + \frac{1}{\sigma^3} \Sigma(x_\alpha - a - bx_{\alpha+l}). \end{aligned}
$$

The solutions of the equations obtained by setting the above derivatives equal to zero are:

$$
(2.6)\qquad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \Sigma(x_\alpha - \hat{a} - \hat{b}x_{\alpha+l})^2 \\ \hat{a} &= \frac{1}{n} \Sigma x_\alpha(1 - \hat{b}) \\ \hat{b} &= \frac{n\Sigma x_\alpha x_{\alpha+l} - (\Sigma x_\alpha)^2}{n\Sigma x_\alpha^2 - (\Sigma x_\alpha)^2}. \end{aligned}
$$

If we now maximize L over the space ω we obtain

$$
(2.7)\qquad \begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \Sigma(x_\alpha - \hat{a})^2 \\ \hat{a} &= \frac{1}{n} \Sigma x_\alpha \end{aligned}
$$

so that we will have

$$
(2.8)\qquad dP_n(\Omega \max) = [2\pi(1 - \hat{b}^2)\Sigma(x_\alpha - \bar{x})^2]^{-\frac{1}{2}n} e^{-\frac{1}{2}n},
$$

$$
(2.9)\qquad dP_n(\omega \max) = [2\pi\Sigma(x_\alpha - \bar{x})^2]^{-\frac{1}{2}n} e^{-\frac{1}{2}n},
$$

$$
(2.10)\qquad \lambda_1 = (1 - \hat{b}^2)^{\frac{1}{2}n},
$$

where b is defined as in (2.6) above. If we set $a = 0$ in (2.1) we may follow a similar procedure and find the criterion ${}_0\lambda_1 = (1 - \hat{b}_0^2)^{\frac{1}{2}n}$ for testing the hypothesis that $b = 0$ if it is known that the population mean equal zero. Here

$$\hat{b}_0 = \frac{\Sigma x_\alpha x_{\alpha+l}}{\Sigma x_\alpha^2}. \tag{2.11}$$

We notice that $\hat{b}$ is the criterion chosen by R. L. Anderson as a measure of serial correlation. He has obtained the distribution of $\hat{b}$ and has computed a number of significance values from this distribution. The distribution is a function of n and l, and Anderson points out that the larger the values of n the smaller the extent to which the significance values depend on l.

In the next section we shall find a distribution which approximates very closely the exact distribution of $\hat{b}$ and which is independent of the lag l.

3. Moments of the likelihood criteria. We shall determine the moments of $\hat{b}_0$ and ${}_0\lambda_1$ when the hypothesis ${}_0H_1$ is true and the moments of $\hat{b}$ and λ_1 when the hypothesis H_1 is true. Let us first consider $\hat{b}_0 = \Sigma x_\alpha x_{\alpha+l}/\Sigma x_\alpha^2$, the criterion we obtained for testing the hypothesis ${}_0H_1$. The moment generating function for the joint distribution of $C_0 = \Sigma x_\alpha x_{\alpha+l}/\sigma^2$ and $V_0 = \Sigma x_\alpha^2/\sigma^2$ is

$$\begin{aligned}\varphi(t_1, t_2) &= E[\exp(C_0 t_1 + V_0 t_2)] \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int \exp\left(C_0 t_1 + V_0 t_2 - \frac{1}{2\sigma^2}\Sigma x_\alpha^2\right) \prod_{\alpha=1}^{n} dx_\alpha.\end{aligned} \tag{3.1}$$

By reference to this last expression it can be seen that

$$E(\hat{b}_0) = \int_{-\infty}^{0} \left.\frac{\partial\varphi(t_1, t_2)}{\partial t_1}\right]_{t_1=0} dt_2, \tag{3.2}$$

and further

$$E(\hat{b}_0^k) = \int_{-\infty}^{0} \cdots \int \left.\frac{\partial^k \varphi(t_1, t_2)}{\partial t_1^k}\right]_{t_1=0} \prod_{j=1}^{k} dt_{2j}, \tag{3.3}$$

in which we set $t_2 = \sum_{j=1}^{k} t_{2j}$.

Now if we write (3.1) as follows:

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int e^{-\frac{1}{2\sigma^2}\sum_{ij} A_{ij}x_ix_j} \Pi\, dx_i, \tag{3.4}$$

we see that $\varphi(t_1, t_2) = |A_{ij}|^{-\frac{1}{2}}$ and $A_{ij} = A_{i+a,j+a}$; that is, this determinant is a circulant. Let us write $A_{ij} = a_h$ where h equal $j - i + 1$ or $j - i + 1 + n$ taking that subscript which gives $1 \le h \le n$, so that we have

$$\varphi^{-2}(t_1, t_2) = \begin{vmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{vmatrix}, \tag{3.5}$$

which expanded by the method of circulants becomes

$$\prod_{k=1}^{n} \sum_{i=1}^{n} a_i \omega_k^{i-1},$$

where the ω_k are the nth roots of unity.

First we shall consider $\hat{b}_0$ (lag 1). Here $a_1 = 1 - 2t_2$, $a_2 = a_n = -t_1$ and a_3 to a_{n-1} equal zero.

$$(3.6)\qquad \begin{aligned} {}_0\varphi_1^{-2}(t_1, t_2) &= \prod_{k=1}^{n} (1 - 2t_2 - t_1(\omega_k + \omega_k^{-1})) \\ &= \prod_{k=1}^{n} \left(1 - 2t_2 - 2t_1 \cos \frac{2\pi k}{n}\right). \end{aligned}$$

For lag l, $a_1 = 1 - 2t_2$, $a_{l+1} = a_{n-l+1} = -t_1$ and the remaining a's $= 0$.

$$(3.7)\qquad \begin{aligned} {}_0\varphi_1^{-2}(t_1, t_2) &= \prod_{k=1}^{n} (1 - 2t_2 - t_1(\omega_k^{l} + \omega_k^{-l})) \\ &= \prod_{k=1}^{n} \left(1 - 2t_2 - 2t_1 \cos \frac{2\pi lk}{n}\right). \end{aligned}$$

We shall develop an approximation to these functions (3.6) and (3.7) as follows:

$$(3.8)\qquad {}_0\varphi_1(t_1, t_2) = \prod_{k=1}^{n} (A + B \cos \alpha_k)^{-\frac{1}{2}} = e^{-\frac{1}{2} \sum_{k=1}^{n} \log (A + B \cos \alpha_k)},$$

where in this case $A = 1 - 2t_2$, $B = -2t_1$ and $\alpha_k = 2\pi lk/n$. Let us now alter the form of this exponent and replace the sum of a finite number of terms involving α_k by an integral of a continuous variable α.

$$(3.9)\qquad {}_0\varphi_1(t_1, t_2) = \exp\left(\frac{-n}{4\pi l} \frac{2\pi l}{n} \sum_{k=1}^{n} \log (A + B \cos \alpha_k)\right).$$

Let us write $2\pi l/n = \Delta\alpha_k$, then we shall have

$$(3.10)\qquad {}_0\varphi_1(t_1, t_2) = \exp\left(\frac{-n}{4\pi l} \sum_{k=1}^{n} \log (A + B \cos \alpha_k)\Delta\alpha_k\right).$$

If we take $B < A$ we see that $A + B \cos \alpha_k$ is never negative; therefore as we let n increase the summation will approach the value of the integral: $\int_0^{2\pi l} \log (A + B \cos \alpha)\, d\alpha$. Let us then replace this summation by its limiting integral. The resulting function will then be an approximation for n large enough. We shall obtain then

$$(3.11)\qquad {}_0\varphi_1(t_1, t_2) \sim \exp\left(\frac{-n}{4\pi l} \int_0^{2\pi l} \log (A + B \cos \alpha)\, d\alpha\right) \qquad B < A,$$

from which by the use of the integral[1] we obtain

$$(3.12)\qquad {}_0\varphi_1(t_1, t_2) \sim \exp\left(-\tfrac{1}{2}n \log \tfrac{1}{2}(A + \sqrt{A^2 - B^2})\right).$$

[1] This integral may be verified by differentiation with respect to the parameter a. It may be found as formula 523 in Pierce's "Short Table of Integrals."

So that

$$ {}_0\varphi_1(t_1, t_2) \sim [\tfrac{1}{2}(A + \sqrt{A^2 - B^2})]^{-\frac{1}{2}n}. \tag{3.13}$$

We now have $\varphi(t_1, t_2)$ represented approximately by a power of a single quantity. The question of how good this approximation is will be discussed in a later paragraph. This is similar to the approximation used by T. Koopmans in the distribution of $\hat{b}_0$. He makes the substitution "to obtain what intuitively seems to be in some sense the closest approximation." He approximates $\prod_{t=1}^{T} (\kappa - \bar{\kappa}_t)^{-\frac{1}{2}}$ where $\bar{\kappa}_t = \cos \dfrac{2\pi t}{T}$ by the process followed in (3.8)–(3.13) in order to find the distribution given in (1.3).

To obtain the corresponding function for

$$ \hat{b} = \frac{\Sigma x_\alpha x_{\alpha+l} - (\Sigma x_\alpha)^2/n}{\Sigma x_\alpha^2 - (\Sigma x_\alpha)^2/n}, \tag{3.14}$$

we follow the same procedure as above for $\hat{b}_0$. Here

$$C = [\Sigma x_\alpha x_{\alpha+l} - (\Sigma x_\alpha)^2/n]/\sigma^2,$$

$$V = [\Sigma x_\alpha^2 - (\Sigma x_\alpha)^2/n]/\sigma^2,$$

and in (3.5) $a_1 = 1 - 2t_2 + 2(t_1 + t_2)/n$; $a_2 = a_n = -t_1 + 2(t_1 + t_2)/n$; and all the other a's $= 2(t_1 + t_2)/n$ so that the expansion of this circulant becomes

$$ \begin{aligned} \varphi_1^{-2}(t_1, t_2) = \prod_{k=1}^{n} \Big\{ &[1 - 2t_2 + 2(t_1 + t_2)/n] + [-t_1 + 2(t_1 + t_2)/n](\omega_k + \omega_k^{-1}) \\ &+ [2(t_1 + t_2)/n] \sum_{i=3}^{n-1} \omega_k^{i-1} \Big\}, \end{aligned} \tag{3.15}$$

and since

$$ \sum_{i=3}^{n-1} \omega_k^{i-1} = \begin{cases} -(\omega_k + \omega_k^{-1} + 1) & k \neq n, \\ n - 3 & k = n, \end{cases} $$

we get

$$ \prod_{k=1}^{n-1} \{1 - 2t_2 - t_1(\omega_k + \omega_k^{-1})\}, \tag{3.16}$$

and for lag l we get

$$ \prod_{k=1}^{n-1} \{1 - 2t_2 - t_1(\omega_k^{l} + \omega_k^{-l})\}. \tag{3.17}$$

These two results are the same as those obtained previously except that the final term, $1 - 2t_1 - 2t_2 = A + B$, of the products is missing. We will then obtain as an approximation to these finite products $\varphi_1^{-2} = {}_0\varphi_1^{-2}/(A + B)$ or

$$ \varphi_1(t_1, t_2) = [\tfrac{1}{2}(A + \sqrt{A^2 - B^2})]^{-\frac{1}{2}n}\sqrt{A + B} \tag{3.18}$$

where $A = 1 - 2t_2$ and $B = -2t_1$.

A method of finding the moments of $\hat{b}_0$ and $\hat{b}$ was outlined in (3.1), (3.2) and (3.3) above. If we perform these operations on ${}_0\varphi_1(t_1, t_2)$ as defined in (3.13) we find

$$\begin{aligned} {}_0\varphi_1 &= u^{-\frac{1}{2}n} \\ \frac{\partial \varphi}{\partial t_1} &= -\tfrac{1}{2}nu^{-\frac{1}{2}n-1}\frac{\partial u}{\partial t_1} \qquad (3.19) \\ \frac{\partial^2 \varphi}{\partial t_1^2} &= -\tfrac{1}{2}nu^{-\frac{1}{2}n-2}\left[\left(-\frac{n}{2}-1\right)\left(\frac{\partial u}{\partial t_1}\right)^2 + u\frac{\partial^2 u}{\partial t_1^2}\right] \end{aligned}$$

where $u|_0 = 1 - 2t_2$, $\left.\frac{\partial u}{\partial t_1}\right|_0 = 0$, $\left.\frac{\partial^2 u}{\partial t_1^2}\right|_0 = \frac{-2}{1-2t_2}$, etc., and the zero subscript indicates that t_1 has been set equal to zero after differentiation. If these values are substituted and the required number of integrations with respect to t_2 are performed, we find the moments of the criterion $\hat{b}_0$ when ${}_0H_1$ is true.

$$\begin{aligned} &M_1 = 0 \qquad M_2 = \frac{1}{n+2} \\ &M_3 = 0 \qquad M_4 = \frac{3}{(n+2)(n+4)} \\ (3.20) \qquad &M_5 = 0 \qquad M_6 = \frac{15}{(n+2)(n+4)(n+6)} \\ &\qquad\qquad \text{etc., or} \\ &M_{2k-1} = 0 \\ &M_{2k} = \frac{1\cdot 3\cdot 5 \cdots (2k-1)}{(n+2)(n+4)\cdots(n+2k)}. \end{aligned}$$

M_{2k} may be verified by the use of an expansion of the generating function (3.13) by a method of Laplace [10]. He gives the expansion of u^{-i} where u is given by the equation $u^2 - 2u + e^2 = 0$ as follows:

$$u^{-i} = \frac{1}{2^i} + \frac{ie^2}{2^{i+2}} + \frac{i(i+3)e^4}{1\cdot 2\cdot 2^{i+4}} + \frac{i(i+4)(i+5)e^6}{1\cdot 2\cdot 3\cdot 2^{i+6}} + \cdots + \frac{i(i+k+1)\cdots(i+2k-1)e^{2k}}{(k-1)!\cdot 2^{i+2k}} + \cdots.$$

We see that $u = 1 + \sqrt{1-e^2}$, and if we set $e = t_1/(\frac{1}{2} - t_2)$ and $i = \frac{1}{2}n$. We obtain ${}_0\varphi_1 = u^{-\frac{1}{2}n}$ as a series in the even powers of t_1. From this we can see that the odd moments are zero and from the form of the coefficients we can verify M_{2k}.

These moments are not contained in the works of the other authors writing on this subject. Although these moments are obtained from an approximate generating function they are, as will be shown later, the exact, not approximate,

moments for $k < n$, for lag 1 and are the exact moments for $k < n/\alpha$ for any lag l where α is the largest common factor of n and the lag l.

To obtain the moments of $\hat{b}$ we follow a similar procedure with $\varphi_1 = u^{-\frac{1}{2}n}(1 - 2t_1 - 2t_2)^{\frac{1}{2}}$. Differentiating φ_1 the required number of times with respect to t_1 and integrating an equal number of times with respect to t_2 gives the following moments:

$$
\begin{aligned}
&M_1 = \frac{-1}{n-1} \qquad\qquad M_2 = \frac{1}{n+1}\\
&M_3 = \frac{-3}{(n-1)(n+3)} \qquad M_4 = \frac{3}{(n+1)(n+3)}\\
&\vdots\\
&M_{2k-1} = \frac{-1\cdot 3\cdot 5 \cdots (2k-1)}{(n-1)(n+3)(n+5)\cdots(n+2k-1)}\\
&M_{2k} = \frac{1\cdot 3\cdot 5 \cdots (2k-1)}{(n+1)(n+3)\cdots(n+2k-1)}.
\end{aligned}
\tag{3.21}
$$

Examination of the moments of $\hat{b}_0$ will show that $\hat{b}_0 = x$ is distributed according to the law

$$K_1(1 - x^2)^{\frac{1}{2}(n-1)} = K_1(1 + x)^{\frac{1}{2}(n-1)}(1 - x)^{\frac{1}{2}(n-1)} \tag{3.22}$$

up to n moments. This distribution is symmetric and we may wish a normal approximation to this curve. The mean is zero and the variance is $1/(n + 2)$. The λ criterion ${}_0\lambda_1^{2/n} = (1 - \hat{b}_0^2) = y$ is distributed according to the law

$$K_2(1 - y)^{-\frac{1}{2}}y^{\frac{1}{2}(n-1)} \tag{3.23}$$

up to $\frac{1}{2}n$ moments. Here the mean is $\dfrac{n+1}{n+2}$ and the variance is $\dfrac{2(n+1)}{(n+2)^2(n+4)}$. If we inspect the moments of $\hat{b}$ we see that the distribution of $\lambda_1^{2/n} = (1 - \hat{b}^2) = z$ follows the law

$$K_3(1 - z)^{-\frac{1}{2}}z^{\frac{1}{2}(n-2)} \tag{3.24}$$

up to $\frac{1}{2}n$ moments, which is the same as the distribution immediately preceding except that n is replaced by $n - 1$. The distributions (3.22), (3.23), and (3.24) are the same for lag l except that the fit is up to n/α, $n/2\alpha$, and $n/2\alpha$ moments respectively where α is the largest common factor of l and n. These restrictions are necessary since the moments as given in (3.20) and (3.21) are obtained from the approximate generating functions (3.13) and (3.18). The exact generating function is given in (4.8) for lag 1 and it is found that the nth or higher derivatives bring contributions from the part of the generating function which was neglected in approximating the generating function. The additional restriction for lag $l \neq 1$ will be seen in the last two paragraphs of section 4. The extra

factor $\frac{1}{2}$ in the second and third case above is due to the fact that only the even moments of (3.20) and (3.21) are used.

We have in (3.23) and (3.24), then, very close approximations to the distributions for the two λ criteria for testing serial effects.

The following table is a comparison of the exact and approximate 5 per cent and 1 per cent points for the distribution of $\hat{b}$. The exact values are taken from the table given by R. L. Anderson. The normal approximation as given by Anderson in his table does not show such close agreement since he used an asymptotic second moment. He indicated that the exact values would have to be used for values of $n < 75$ in place of the values from the normal approximation which he obtained. Here we see that the normal approximation may be used for n somewhat less than 75. The Pearson Type I approximation was obtained by using the first two moments of $\hat{b}$. The curve obtained is:

$$\frac{(1 + x)^{p-1}(1 - x)^{q-1}}{B(p, q)2^{p+q-1}} \tag{3.25}$$

in which $p = \dfrac{(n - 1)(n - 2)}{2(n - 3)}$ and $q = \dfrac{n(n - 1)}{2(n - 3)}$.

The exact values marked with an asterisk in the table differ slightly from those previously published. They are more precise values from the exact distribution which R. L. Anderson has made available to the author.

Positive tail

N	*Exact*	5% *Type I*	*Normal*	*Exact*	1% *Type I*	*Normal*
5	.253	.317	.281	.297	.527	.501
10	.360	.362	.350	.525	.533	.541
15	.328	.329	.323	.475	.477	.486
20	.299	.299	.296	.432	.433	.440
25	.276	.276	.274	.398	.398	.404
30	.257	.257	.255	.370	.371	.375
45	.218	.218	.217	.313*	.313	.316
75	.174*	.174	.174	.250	.250	.251

Negative tail

N	*Exact*	5% *Type I*	*Normal*	*Exact*	1% *Type I*	*Normal*
5	.753	.742	.781	.798	.858	1.000
10	.564	.562	.572	.705	.702	.763
15	.462	.461	.466	.597	.596	.629
20	.399	.399	.401	.524	.524	.545
25	.356	.356	.357	.473	.473	.487
30	.324*	.324	.324	.433	.433	.444
45	.262	.262	.262	.356	.356	.362
75	.201*	.201	.201	.276	.276	.278

4. Alternative expansions of the generating functions. In this section we shall determine the exact generating functions which were approximated in (3.12) and (3.16) and obtain these same approximations in another manner. This development will enable us to see how good the approximation is in the sense that it gives a certain number of exact moments. The determinant in (3.4) for lag l and mean zero can be written

$$(4.1) \qquad A_n = \begin{vmatrix} a & b & & & & & b \\ b & a & b & & & & \\ & b & a & b & & 0 & \\ & & & \cdots & & & \\ & & & & \cdots & & \\ & & 0 & & b & a & b \\ b & & & & & b & a \end{vmatrix}_n$$

where $a = 1 - 2t_2$, $b = -t_1$ and all the other elements are zero. The b in the upper right corner and lower left corner indicate the value b in the a_{1n} and a_{n1} positions. Let us define the following determinants:

$$(4.2) \qquad B_n = \begin{vmatrix} a & b & & & & & \\ b & a & b & & & & \\ & b & a & b & & 0 & \\ & & & \cdots & & & \\ & & & & \cdots & & \\ & 0 & & & & & \\ & & & & b & a & b \\ b & & & & & b & a \end{vmatrix}_n \qquad C_n = \begin{vmatrix} b & a & b & & & & \\ & b & a & b & & & \\ & & b & a & b & 0 & \\ & & & \cdots & & & \\ & & & & \cdots & & \\ & 0 & & & & & \\ & & & & & b & a \\ b & & & & & & b \end{vmatrix}_n$$

$$D_n = \begin{vmatrix} a & b & & & & & \\ b & a & b & & & & \\ & b & a & b & & 0 & \\ & & & \cdots & & & \\ & & & & \cdots & & \\ & 0 & & & b & a & b \\ & & & & & b & a \end{vmatrix}_n$$

We see that

$$(4.3) \qquad \begin{aligned} A_n &= B_n + (-1)^{n-1} b C_{n-1}, \\ B_n &= D_n + (-1)^{n-1} b^n, \\ C_n &= b^n + (-1)^{n-1} b D_{n-1}, \end{aligned}$$

and A_n can be expressed in terms of D_n by substituting for B_n and C_n in the equation for A_n.

$$(4.4) \qquad A_n = D_n - b^2 D_{n-2} + 2(-1)^{n-1} b^n.$$

We can obtain an expression for D_n if we expand this determinant by the first row. This gives

$$(4.5) \qquad D_n = aD_{n-1} - b^2 D_{n-2}.$$

Since this is a second order difference equation, the solution may be written $D_n = k_1u^n + k_2v^n$ where u, v are roots of the equation $x^2 - ax + b^2 = 0$. Now,

$$
\begin{aligned}
D_1 &= u + v = k_1u + k_2v, \\
D_2 &= u^2 + v^2 + uv = k_1u^2 + k_2v^2
\end{aligned} \tag{4.6}
$$

so that we can determine k_1 and k_2. We now see that

$$
D_n = \frac{u^{n+1} - v^{n+1}}{u - v} \tag{4.7}
$$

which upon substitution in the equation for A_n gives

$$
\begin{aligned}
A_n &= u^n + v^n + 2(-1)^{n-1}b^n, \\
&= u^n + v^n - 2t_1^n,
\end{aligned} \tag{4.8}
$$

where $u, v = \frac{1}{2}(1 - 2t_2) \pm \sqrt{(1 - 2t_2)^2 - 4t_1^2}$. Now ${}_0\varphi_1(t_1, t_2) = A_n^{-\frac{1}{2}}$ and it is easily seen from the form of A_n directly above that derivatives up to the nth order with respect to t_1 in which t_1 is then set equal to zero will be given by derivatives of $A_n = u^n$ and this is the approximation (3.13) found by other methods.

The determinant in (3.4) for lag 1 and mean not equal to zero can be written

$$
A_n = \begin{vmatrix}
a & b & & & & & b \\
b & a & b & & c & & \\
 & b & a & b & & & \\
 & & & \cdots & & & \\
 & & & & \cdots & & \\
 & c & & & & & \\
 & & & & b & a & b \\
b & & & & & b & a
\end{vmatrix}
\qquad
\begin{aligned}
a &= 1 - 2t_2 + 2(t_1 + t_2)/n, \\
b &= -t_1 + 2(t_1 + t_2)/n, \\
c &= 2(t_1 + t_2)/n.
\end{aligned} \tag{4.9}
$$

Let us define the following determinants:

$$
B_n = \begin{vmatrix}
b & a & b & & & & \\
 & b & a & b & & c & \\
 & & b & a & b & & \\
 & & & & \cdots & & \\
 & & & & & \cdots & \\
 & c & & & & & \\
 & & & & & b & a \\
b & & & & & & b
\end{vmatrix}
\qquad
C_n = \begin{vmatrix}
a & b & & & & & \\
b & a & b & & c & & \\
 & b & a & b & & & \\
 & & & \cdots & & & \\
 & & & & \cdots & & \\
 & c & & & & & \\
 & & & & b & a & b \\
b & & & & & b & a
\end{vmatrix}
$$

$$
D_n = \begin{vmatrix}
b & a & b & & & & \\
 & b & a & b & & c & \\
 & & b & a & b & & \\
 & & & & \cdots & & \\
 & & & & & \cdots & \\
 & c & & & & & \\
 & & & & & b & a \\
 & & & & & & b
\end{vmatrix}
\qquad
E_n = \begin{vmatrix}
a & b & & & & & \\
b & a & b & & c & & \\
 & b & a & b & & & \\
 & & & \cdots & & & \\
 & & & & \cdots & & \\
 & c & & & & & \\
 & & & & b & a & b \\
 & & & & & b & a
\end{vmatrix} \tag{4.10}
$$

If we replace the b in the upper right corner of A_n by $c + (b - c)$ we obtain

$$A_n = C_n + (-1)^{n-1}(b - c)B_{n-1}. \tag{4.11}$$

If we replace the b in the lower left corner of B_n and C_n by $c + (b - c)$ we obtain

$$\begin{aligned} B_n &= D_n + (-1)^{n-1}(b - c)E_{n-1}, \\ C_n &= E_n + (-1)^{n-1}(b - c)D_{n-1}. \end{aligned} \tag{4.12}$$

We now have A_n in terms of D_n and E_n. We must now evaluate D_n and E_n.

$$D_n = \begin{vmatrix} 1 & 1 & 1 & \cdots & & & 1 \\ 0 & b & a & b & & & \\ 0 & & b & a & b & & c \\ \cdot & & & \cdot & \cdot & \cdot & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & c & & & & & \\ & & & & & b & a \\ 0 & & & & & & b \end{vmatrix}_{n+1} = \begin{vmatrix} 1 & 1 & 1 & \cdots & & & 1 \\ -c & r & s & r & & & \\ -c & & r & s & r & & 0 \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & \cdot \\ & & 0 & & & & \\ & & & & & r & s \\ -c & & & & & & r \end{vmatrix}_{n+1} \tag{4.13}$$

where $r = b - c$ and $s = a - c$ and the second determinant above is obtained from the first by subtracting c times the first row from all other rows. Writing this last determinant as the sum of two determinants by separating the first column we get

$$D_n = r^n - cF_{n+1}. \tag{4.14}$$

Combining the above difference equations we obtain

$$A_n = E_n - r^2E_{n-2} + 2(-1)^n rcF_n + 2(-1)^{n-1}r^n \tag{4.15}$$

and see that we must obtain E_n and F_n.

Expansion of F_{n+1} by the second column gives

$$F_{n+1} = -G_n + rF_n \tag{4.16}$$

and expanding G_n by the last row we get

$$G_n = rG_{n-1} + (-1)^{n-1}H_{n-1}. \tag{4.17}$$

$$F_{n+1} = \begin{vmatrix} 0 & 1 & 1 & \cdots & & & 1 \\ 1 & r & s & r & & 0 & \\ 1 & & r & s & r & & \\ \cdot & & & \cdot & \cdot & \cdot & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & & 0 & & & & \\ & & & & & r & s \\ 1 & & & & & & r \end{vmatrix}_{n+1} \qquad G_n = \begin{vmatrix} 1 & s & r & & & & \\ 1 & r & s & r & & 0 & \\ 1 & & r & s & r & & \\ \cdot & & & \cdot & \cdot & \cdot & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & & 0 & & & & \\ & & & & & r & s \\ 1 & & & & & & r \end{vmatrix}_{n} \tag{4.18}$$

$$H_n = \begin{vmatrix} s & r & & & & \\ r & s & r & & 0 & \\ & r & s & r & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \cdot \\ 0 & & & & & \\ & & & r & s & r \\ & & & & r & s \end{vmatrix}$$

H_n is the same type as (4.1), therefore $H_n = \dfrac{u^{n+1} - v^{n+1}}{u - v}$ where u and v are the roots of the equation $x^2 - sx + r^2 = 0$, so that (4.17) becomes

$$G_n - rG_{n-1} = (-1)^{n-1}\frac{u^n - v^n}{u - v} \tag{4.19}$$

and the solution of this equation gives

$$G_n = \frac{r^n}{2r + s} + \frac{(-1)^{n-1}[r(u^n - v^n) + u^{n+1} - v^{n+1}]}{(u - v)(2r + s)}. \tag{4.20}$$

Introducing this expression into (4.16) we find

$$F_n = (-1)^{n-1}\frac{u^n - v^n}{(u - v)(2r + s)} - \frac{nr^{n-1}}{2r + s} \tag{4.21}$$

$$E_n = \begin{vmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 0 & a & b & & & & \\ 0 & b & a & b & & c & \\ \cdot & & \cdot & \cdot & \cdot & & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & & c & & & & \\ & & & & b & a & b \\ 0 & & & & & b & a \end{vmatrix}_{n+1} = \begin{vmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ -c & s & r & & & & \\ -c & r & s & r & & 0 & \\ \cdot & & \cdot & \cdot & \cdot & & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & & 0 & & & & \\ & & & & r & s & r \\ -c & & & & & r & s \end{vmatrix}_{n+1} \tag{4.22}$$

where the second determinant is obtained from the first by subtracting c times the first row from all other rows. If we separate this last determinant on the first column we get

$$E_n = H_n - cI_{n+1} \tag{4.23}$$

$$I_n = \begin{vmatrix} 0 & 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & s & r & & & & \\ 1 & r & s & r & & 0 & \\ \cdot & & \cdot & \cdot & \cdot & & \\ \cdot & & & & \cdot & \cdot & \cdot \\ & & 0 & & & & \\ & & & & r & s & r \\ 1 & & & & & r & s \end{vmatrix} \qquad J_n = \begin{vmatrix} 0 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 \\ 1 & s & r & & & & & \\ 1 & r & s & r & & 0 & & \\ \cdot & & \cdot & \cdot & \cdot & & & \\ \cdot & & & & \cdot & \cdot & \cdot & \\ & & 0 & & & r & s & r \\ & & & & & & r & s \\ 1 & & & & & & r & r \end{vmatrix} \tag{4.24}$$

Expanding I_n by the last row, we get

$$I_n = (-1)^{n-1}G_{n-1} - rJ_{n-1} + sI_{n-1}. \tag{4.25}$$

Expanding J_n by the last column, we get

$$J_n = (-1)^{n-1}G_{n-1} + rI_{n-1}. \tag{4.26}$$

If we combine these last two equations we find

$$I_n - sI_{n-1} + r^2I_{n-2} = (-1)^{n-1}(G_{n-1} + rG_{n-2}). \tag{4.27}$$

If we now solve this difference equation for I_n, substitute this solution in the equation for E_n, and in turn substitute this result and the expression we obtained for F_n in (4.15), we get

$$
\begin{aligned}
A_n &= \frac{u^n + v^n + 2(-1)^{n-1} r^n}{2r + s} \\
&= \frac{u^n + v^n - 2t_1^n}{1 - 2t_1 - 2t_2}.
\end{aligned} \tag{4.28}
$$

The final form results since $2r + s = 1 - 2t_1 - 2t_2$, $r = -t_1$. u and v have the same values as before. If we compare this result with that obtained in (4.8) for mean equal to zero we see that this is the same except that here we have the added factor, $1 - 2t_1 - 2t_2$, in the denominator. We have a similar result then for the approximation for derivatives of $\varphi_1(t_1, t_2) = A_n^{-\frac{1}{2}}$ for $t_1 = 0$. Here this approximation is $A_n = u^n/(1 - 2t_1 - 2t_2)$, the same result as that obtained in (3.18). This approximation will yield moments which are exact for $n > \alpha k$ for any lag l where α is the largest common factor of n and l. The reason for this restriction is easily seen if we consider the expansion obtained in (3.7), for

$$
\frac{\partial \varphi(t_1, t_2)}{\partial t_1} = -\tfrac{1}{2}\varphi(t_1, t_2) \sum_{k=1}^{n} \frac{-2 \cos \dfrac{2\pi l k}{n}}{1 - 2t_2 - 2t_1 \cos \dfrac{2\pi l k}{n}} \tag{4.29}
$$

with $t_1 = 0$,

$$
\begin{aligned}
\left.\frac{\partial \varphi(t_1, t_2)}{\partial t_1}\right]_{t_1=0} &= -\tfrac{1}{2}\varphi(0, t_2) \sum_{k=1}^{n} \frac{-2 \cos \dfrac{2\pi l k}{n}}{1 - 2t_2} \\
&= \frac{\varphi(0, t_2)}{1 - 2t_2} \sum_{k=1}^{n} \cos \frac{2\pi l k}{n}.
\end{aligned} \tag{4.30}
$$

Further

$$
\left.\frac{\partial^2 \varphi(t_1, t_2)}{\partial t_1^2}\right]_{t_1=0} = \frac{\varphi(0, t_2)}{(1 - 2t_2)^2} \left[\left(\sum_{k=1}^{n} \cos \frac{2\pi l k}{n} \right)^2 + 2 \sum_{k=1}^{n} \cos^2 \frac{2\pi l k}{n} \right] \tag{4.31}
$$

and the mth partial derivative will contain the sum of the mth powers of the cosines. These are the sums of the powers of the real parts of the roots of unity and it is easily seen that $\Sigma \cos^m \dfrac{2\pi l k}{n} = \Sigma \cos^m \dfrac{2\pi k}{n}$ only for $m < \alpha k$ where α is the largest common factor of n and l.

To change the moment generating function of $\hat{b}_0$ to that of $\hat{b}$ we must drop the last term of the product. In the above expressions we then have $\sum_{k=1}^{n-1} \cos^m \dfrac{2\pi l k}{n}$ and the same conclusion will hold.

5. Application to successive differences. If we change slightly the function $\eta = \delta_{n-1}^2/V_n$ investigated by von Neumann and Williams we find the moments and distribution greatly simplified. Let us define

$$\delta_n^2 = \sum_{i=1}^{n} (x_i - x_{i+1})^2 \tag{5.1}$$

where $x_{n+1} = x_1$ and consider the ratio ${}_0\eta_1$ of δ_n^2 to Σx_i^2. Now,

$$\delta_n^2 = 2\Sigma x_i^2 - 2\Sigma x_i x_{i+1} \tag{5.2}$$

therefore

$$ {}_0\eta_1 = \frac{\delta_n^2}{\Sigma x_i^2} = 2(1 - \hat{b}_0) \tag{5.3}$$

and we may find the moments and distribution of ${}_0\eta_1$ directly from those of $\hat{b}_0$. We find the moments to be:

$$\begin{aligned} m_1 &= 2 & m_2 &= \frac{2^2(n+3)}{n+2} \\ m_3 &= \frac{2^3(n+5)}{n+2} & m_4 &= \frac{2^4(n+5)(n+7)}{(n+2)(n+4)} \end{aligned} \tag{5.4}$$

$$m_k = \frac{2^k(n+2k-1)!}{(n+k)!\,(n+2)(n+4)\cdots(n+2k-2)}, \qquad (k < n)$$

and the ratio ${}_0\eta_1$ is distributed according to the law

$$C_{10}\eta_1^{\frac{1}{2}(n-1)}(4 - {}_0\eta_1)^{\frac{1}{2}(n-1)} \tag{5.5}$$

up to n moments.

If we replace x_i in the above ratios by $x_i - \bar{x}$ we find the moments of the ratio $\eta_1 = \delta_n^2/\Sigma(x_i - \bar{x})^2$ to be:

$$\begin{aligned} m_1 &= \frac{2n}{n-1} & m_2 &= \frac{2^2\,n(n+3)}{(n-1)(n+1)} \\ m_3 &= \frac{2^3\,n(n+4)(n+5)}{(n-1)(n+1)(n+3)} & m_4 &= \frac{2^4\,n(n+5)(n+6)(n+7)}{(n-1)(n+1)(n+3)(n+5)} \end{aligned} \tag{5.6}$$

$$m_k = \frac{2^k\,n(n+2k-1)!}{(n+k)!\,(n-1)(n+1)(n+3)\cdots(n+2k-3)}$$

and $(\eta_1 - 2)^2 = z$ has the distribution

$$C_2 z^{-\frac{1}{2}}(4 - z)^{\frac{1}{2}(n-2)} \tag{5.7}$$

up to $\frac{1}{2}n$ moments.

The ratio $\delta_n^2/\Sigma x_i^2$ compares the variation of the first differences to that of the original variates. We might wish to compare the variation of the second

differences to that of the first differences. For this purpose let us form the ratio

$$\eta_2 = \frac{\sum_{i=1}^{n} (x_i - 2x_{i+1} + x_{i+2})^2}{\sum_{i=1}^{n} (x_i - x_{i+1})^2} \qquad \begin{aligned} x_{n+1} &= x_1 \\ x_{n+2} &= x_2 \end{aligned} \tag{5.8}$$

to test the hypothesis H_{η_2} that the variation of the second differences compared to the variation of the first differences is such as would occur by chance. Let $x_1, x_2, \cdots, x_n$ be normally distributed with mean a and variance σ^2. The ratio η_2 is independent of the mean value of the variates, therefore we may consider a distribution with mean equal zero. We shall develop the mean and variance of η_2 when the hypothesis to be tested is true. The moment generating function for the joint distribution of $D_2 = \Sigma(x_i - 2x_{i+1} + x_{i+2})^2/2\sigma^2$ and $D_1 = \Sigma(x_i - x_{i+1})^2/2\sigma^2$ is

$$\begin{aligned} \varphi(t_1, t_2) &= E[\exp (D_2 t_1 + D_1 t_2)] \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int \exp\left(D_2 t_1 + D_1 t_2 - \frac{1}{2\sigma^2}\Sigma x_i^2\right) \prod_{i=1}^{n} dx_i . \end{aligned} \tag{5.9}$$

We may find the moments then by a process similar to that outlined in (3.2) and (3.3). The next few steps are identical with (3.4) and (3.5). For the present problem, however, $a_1 = 1 - 6t_1 - 2t_2$, $a_2 = 4t_1 + t_2$, $a_3 = -t_1$ so that

$$\begin{aligned} \varphi^{-2}(t_1, t_2) &= \prod_{k=1}^{n} [1 - 6t_1 - 2t_2 + (4t_1 + t_2)(\omega_k^1 + \omega_k^{-1}) - t_1(\omega_k^2 + \omega_k^{-2})], \\ &= \prod_{k=1}^{n} \left[1 - 6t_1 - 2t_2 + (8t_1 + 2t_2) \cos \frac{2\pi k}{n} - 2t_1 \cos \frac{4\pi k}{n}\right], \\ &= \prod_{k=1}^{n} \left[a + b \cos \frac{2\pi k}{n} + c \cos \frac{4\pi k}{n}\right]. \end{aligned} \tag{5.10}$$

If we follow the same procedure indicated in (3.8) to (3.13) we obtain successively

$$\varphi(t_1, t_2) = \prod_{k=1}^{n} (a + b \cos \alpha_k + c \cos 2\alpha_k)^{-\frac{1}{2}} \tag{5.11}$$

$$= e^{-\frac{1}{2}\sum_{k=1}^{n} \log (a+b\cos\alpha_k + c\cos 2\alpha_k)} \tag{5.12}$$

$$= \exp\left(\frac{-n}{4\pi}\frac{2\pi}{n}\sum_{k=1}^{n} \log (a + b \cos \alpha_k + c \cos 2\alpha_k)\right) \tag{5.13}$$

and replace the summation by the integral which is the limit of the summation as $n \to \infty$.

$$\varphi(t_1, t_2) \sim \exp\left(\frac{-n}{4\pi}\int_0^{2\pi} \log (a + b \cos \alpha + c \cos 2\alpha)\, d\alpha\right) \tag{5.14}$$

$$\sim \exp\left(-\frac{n}{2}\log\left[\kappa\, \frac{1 + \sqrt{1 - \delta^2}}{2}\, \frac{1 + \sqrt{1 - \eta^2}}{2}\right]\right) \tag{5.15}$$

where $\kappa = a - c; \quad \eta, \delta = \dfrac{b \pm \sqrt{b^2 + 8c^2 - 8ac}}{2(a - c)}.$

We then have approximately

$$(5.16) \qquad \varphi(t_1, t_2) \sim [\tfrac{1}{4}\kappa(1 + \sqrt{1 - \delta^2})(1 + \sqrt{1 - \eta^2})]^{-\frac{1}{2}n}.$$

(5.14) follows from (5.13) if we replace the summation by an integral, and (5.15) is obtained in the following manner: replace $\cos 2\alpha$ by $2 \cos^2 \alpha - 1$ and factor the resulting quadratic and integrate the factors separately.

$$\int_0^{2\pi} \log (a + b \cos \alpha + c \cos 2\alpha)\, d\alpha = \int_0^{2\pi} \log (a - c + b \cos \alpha + 2c \cos^2 \alpha)\, d\alpha$$

$$(5.17) \qquad = \int_0^{2\pi} \log \kappa \, d\alpha + \int_0^{2\pi} \log (1 + \delta \cos \alpha)\, d\alpha + \int_0^{2\pi} \log (1 + \eta \cos \alpha)\, d\alpha$$

$$= 2\pi \log \kappa + 2\pi \log \tfrac{1}{2}(1 + \sqrt{1 - \delta^2}) + 2\pi \log \tfrac{1}{2}(1 + \sqrt{1 - \eta^2}).$$

If we now expand (5.16) by multiplying the factors within the brackets and substitute for κ, η and δ we find

$$(5.18) \qquad \varphi(t_1, t_2) \sim [A + B + C + D]^{-\frac{1}{2}n} = P^{-\frac{1}{2}n},$$

where

$$(5.19) \qquad \begin{aligned} a &= 1 - 6t_1 - 2t_2, & A &= \tfrac{1}{4}(1 - 4t_1 - 2t_2), \\ b &= 8t_1 + 2t_2, & B &= \tfrac{1}{4}[1 - 12t_1 - 4t_2 + 8t_1t_2 + 2t_2^2 \\ & & &\qquad - 2(4t_1 + t_2)\sqrt{t_2^2 + 4t_1}]^{\frac{1}{2}}, \\ c &= -2t_1, & C &= \tfrac{1}{4}[1 - 12t_1 - 4t_2 + 8t_1t_2 + 2t_2^2 \\ & & &\qquad + 2(4t_1 + t_2)\sqrt{t_2^2 + 4t_1}]^{\frac{1}{2}}, \\ & & D &= \tfrac{1}{4}(1 - 16t_1 - 4t_2)^{\frac{1}{2}}. \end{aligned}$$

From (5.18) $P = A + B + C + D$ and at $t_1 = 0$

$$(5.20) \qquad \begin{aligned} P &= \tfrac{1}{4}(1 + \sqrt{1 - 4t_2})^2, \\ \frac{\partial P}{\partial t_1} &= -2[1 + 2(1 - 4t_2)^{-\frac{1}{2}}], \\ \frac{\partial^2 P}{\partial t_1^2} &= \frac{-32}{(1 - 4t_2)^{\frac{3}{2}}} - \frac{(1 - \sqrt{1 - 4t_2})^2}{2t_2^2}. \end{aligned}$$

Now

$$(5.21) \qquad \begin{aligned} \frac{\partial \varphi}{\partial t_1} &= -\tfrac{1}{2}nP^{-\frac{1}{2}n-1}\frac{\partial P}{\partial t_1} \\ \frac{\partial^2 \varphi}{\partial t_1^2} &= -\tfrac{1}{2}nP^{-\frac{1}{2}n-2}\left[(-\tfrac{1}{2}n - 1)\left(\frac{\partial P}{\partial t_1}\right)^2 + P\frac{\partial^2 P}{\partial t_1^2}\right]. \end{aligned}$$

If we substitute in this formula and integrate the first with respect to t_2 we shall obtain the first moment of the ratio η_2. If we integrate the second twice with respect to t_2, we shall obtain the second moment of the ratio η_2. We find these moments to be

$$(5.22)\qquad \begin{gathered} M_1 = \frac{3n+2}{n+1} \qquad M_2 = \frac{9n^2+23n+12}{(n+1)(n+2)} \\ \sigma^2 = \frac{2n^2+7n+4}{(n+1)^2(n+2)}. \end{gathered}$$

6. Likelihood criteria for multiple serial correlation. Given a sample of n observations, $x_1, x_2, \cdots, x_n$, we shall assume that they are distributed according to the law

$$(6.1)\qquad dP_n = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{\alpha=1}^{n}\left(x_\alpha - a - \sum_{i=1}^{r} b_i x_{\alpha-l_i}\right)^2} dx_1 \cdots dx_n \qquad (x_i = x_{n+i})$$

that is, that each variate, say at the time t, has as its mean value a linear function of the variates at time $t - l_1, t - l_2$, etc. Let us investigate the likelihood criteria for testing the hypothesis, H_r, that each variate is independent of the others; i.e. that the $b_i = 0$ $(i = 1, \cdots, r)$. For the hypothesis H_r we define the space Ω and the space ω, as follows:

$$(6.2)\qquad \begin{cases} \Omega: & \sigma^2 > 0 \quad -\infty < a, b_i < \infty \\ \omega: & \sigma^2 > 0 \quad -\infty < a < \infty, \quad b_i = 0, \end{cases}$$

we find the likelihood ratio criterion

$$(6.3)\qquad \lambda_r^{2/n} = \frac{|a_{ij}|}{a_{00}\,|a_{pq}|} \qquad \begin{matrix} i, j = 0, 1, \cdots, r \\ p, q = 1, \cdots, r \end{matrix}$$

in which

$$(6.4)\qquad \begin{aligned} a_{00} &= \sum_\alpha (x_\alpha - \bar{x})^2 \\ a_{0i} &= \sum_\alpha (x_\alpha - \bar{x})(x_{\alpha+l_i} - \bar{x}) \\ a_{ij} &= \sum_\alpha (x_{\alpha+l_i} - \bar{x})(x_{\alpha+l_j} - \bar{x}) \end{aligned}$$

and it is noted that $a_{ii} = a_{00}$ and if the l_i are equispaced $a_{i,i+c} = a_{0c}$. λ_r is a statistic which measures how completely each variate at time t can be expressed as a linear function of variates spaced at time $t - l_1, t - l_2$, etc.

Next we shall develop a statistic for testing the hypothesis, $H_{r,m}$, that of the set of the values b_i $(i = 1, \cdots, r)$ in (6.1) the subset $b_{m+1}, b_{m+2}, \cdots, b_r = 0$. Here we have the same likelihood function but for $H_{r,m}$, we define the spaces Ω and ω as follows:

$$(6.5)\qquad \begin{cases} \Omega: & \sigma^2 > 0 \quad -\infty < a, b_i < \infty; \\ \omega: & \sigma^2 > 0 \quad -\infty < a, b_u < \infty, \quad b_w = 0, \end{cases}$$

$$u = (1, \cdots, m), \quad w = (m+1, \cdots, r),$$

and obtain the criterion

$$\lambda_{r,m}^{2/n} = \frac{|a_{ij}|\,|a_{uv}|}{|a_{pq}|\,|a_{st}|} \qquad \begin{array}{r} i, j = 0, 1, \cdots, r, \\ p, q = 1, \cdots, r, \\ s, t = 0, 1, \cdots, m, \\ u, v = 1, \cdots, m, \\ (m < r). \end{array} \tag{6.6}$$

The form and the derivation of these λ criteria parallels very closely that of the likelihood ratio criteria obtained in multivariate analysis for testing significance of regression coefficients.

CASE I. If we set $r = 1$ in λ_r we obtain

$$\lambda_1^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}}{a_{00} \quad a_{00}} = 1 - \frac{a_{01}^2}{a_{00}^2} = 1 - \hat{b}^2, \tag{6.7}$$

for which the distribution is given in (3.24).

CASE II. If we set $r = 2$, we have

$$\lambda_2^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{11} & a_{12} \\ a_{02} & a_{12} & a_{22} \end{vmatrix}}{a_{00} \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}}, \tag{6.8}$$

for which if we take $l_1 = 1$, $l_2 = 2$ we get

$$\frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{00} & a_{01} \\ a_{02} & a_{01} & a_{00} \end{vmatrix}}{a_{00} \begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}}. \tag{6.9}$$

The expanded form of this numerator is $a_{00}^3 + 2a_{01}^2a_{02} - a_{00}a_{02}^2 - 2a_{01}^2a_{00}$.

Let us consider

$$\varphi(\theta_0, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int e^{-\frac{1}{2\sigma^2}\{\Sigma x_\alpha^2(1-\theta_0) - \theta_1 \Sigma x_\alpha x_{\alpha+1} - \theta_2 \Sigma x_\alpha x_{\alpha+2}\}} \, \Pi\, dx_\alpha. \tag{6.10}$$

We shall find the mean and variance of ${}_0\lambda_2^{2/n}$ (mean $= 0$) when the hypothesis ${}_0H_r$ ($r = 2$) is true. We can find the first moment of ${}_0\lambda_2^{2/n}$ then by performing the following operations: (a) compute

$$\frac{\partial^3 \varphi}{\partial \theta_0^3} + 2\frac{\partial^2 \varphi}{\partial \theta_1^2}\frac{\partial \varphi}{\partial \theta_2} - \frac{\partial \varphi}{\partial \theta_0}\frac{\partial^2 \varphi}{\partial \theta_2^2} - 2\frac{\partial^2 \varphi}{\partial \theta_1^2}\frac{\partial \varphi}{\partial \theta_0}, \tag{6.11}$$

(this will give the first moment of the numerator) and set $\theta_2 = 0$, (b) integrate from $-\infty$ to $\theta_0 + \theta_1$ with respect to $\theta_0 + \theta_1 = \zeta$, from $-\infty$ to $\theta_0 - \theta_1$, with respect to $\theta_0 - \theta_1 = \xi$, and set $\theta_1 = 0$ (at this point we will have the first moment of the third order determinant divided by the second order determinant), (c)

integrate with respect to θ_0 from $-\infty$ to 0. The reason for step (*b*) is easily seen since the second order determinant $a_{00}^2 - a_{01}^2$ may be written $(a_{00} - a_{01})(a_{00} + a_{01})$.

Further moments may be computed in a similar manner. ${}_0\varphi_2(\theta_0, \theta_1, \theta_2)$ may be written as a determinant in the manner indicated in (3.4) and (3.5). Here, $a_1 = 1 - \theta_0$, $a_2 = a_n = -\frac{1}{2}\theta_1$ and $a_3 = a_{n-1} = -\frac{1}{2}\theta_2$ and a_4 to $a_{n-2} = 0$, then

$$(6.12) \qquad {}_0\varphi_2^{-2}(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^{n} \sum_{i=1}^{n} a_i \omega_k^{i-1} = \prod_{k=1}^{n} \left(a_1 + 2a_2 \cos \frac{2\pi k}{n} + 2a_3 \cos \frac{4\pi k}{n}\right).$$

We shall approximate ${}_0\varphi_2(\theta_0, \theta_1, \theta_2)$ by the method contained between (5.11) and (5.18). We set

$$(6.13) \qquad {}_0\varphi_2(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^{n} \left(a + b \cos \frac{2\pi k}{n} + c \cos \frac{4\pi k}{n}\right)^{-\frac{1}{2}}$$

and obtained

$$(6.14) \qquad {}_0\varphi_2(\theta_0, \theta_1, \theta_2) \sim [A + B + C + D]^{-\frac{1}{2}n} = P^{-\frac{1}{2}n}$$

where

$$(6.15) \qquad \begin{aligned} A &= \tfrac{1}{4}(a - c) \\ B &= \tfrac{1}{8}(4a^2 - 4c^2 - 2b^2 - 2bE^{\frac{1}{2}})^{\frac{1}{2}} = \tfrac{1}{8}\beta^{\frac{1}{2}} \\ C &= \tfrac{1}{8}(4a^2 - 4c^2 - 2b^2 + 2bE^{\frac{1}{2}})^{\frac{1}{2}} = \tfrac{1}{8}\gamma^{\frac{1}{2}} \\ D &= \tfrac{1}{4}((a + c)^2 - b^2)^{\frac{1}{2}} \qquad\qquad = \tfrac{1}{4}\epsilon^{\frac{1}{2}} \\ E &= b^2 + 8c(c - a). \end{aligned}$$

It is easily seen that we may operate (differentiate and integrate) with a, b, c, in place of θ_0, θ_1, θ_2 respectively. Therefore, we compute

$$(6.16) \qquad \begin{aligned} \frac{\partial \varphi}{\partial c} &= -\tfrac{1}{2}nP^{-\frac{1}{2}n-1} \frac{\partial P}{\partial c} \\ \frac{\partial^2 \varphi}{\partial c^2} &= -\tfrac{1}{2}nP^{-\frac{1}{2}n-2}\left[(-\tfrac{1}{2}n - 1)\left(\frac{\partial P}{\partial c}\right)^2 + P\frac{\partial^2 P}{\partial c^2}\right] \end{aligned}$$

and since $P = A + B + C + D$ we compute

$$(6.17) \qquad \begin{aligned} \frac{\partial A}{\partial c} &= -\frac{1}{4}; & \frac{\partial^2 A}{\partial c^2} &= 0 \\ \frac{\partial B}{\partial c} &= \tfrac{1}{16}\beta^{-\frac{1}{2}}\frac{\partial \beta}{\partial c}; & \frac{\partial^2 B}{\partial c^2} &= \frac{1}{16}\left[-\tfrac{1}{2}\beta^{-\frac{3}{2}}\left(\frac{\partial \beta}{\partial c}\right)^2 + \beta^{-\frac{1}{2}}\frac{\partial^2 \beta}{\partial c^2}\right] \\ \frac{\partial C}{\partial c} &= \tfrac{1}{16}\gamma^{-\frac{1}{2}}\frac{\partial \gamma}{\partial c}; & \frac{\partial^2 C}{\partial c^2} &= \frac{1}{16}\left[-\tfrac{1}{2}\gamma^{-\frac{3}{2}}\left(\frac{\partial \gamma}{\partial c}\right)^2 + \gamma^{-\frac{1}{2}}\frac{\partial^2 \gamma}{\partial c^2}\right] \\ \frac{\partial D}{\partial c} &= \tfrac{1}{8}\epsilon^{-\frac{1}{2}}\frac{\partial \epsilon}{\partial c}; & \frac{\partial^2 D}{\partial c^2} &= \frac{1}{8}\left[-\tfrac{1}{2}\epsilon^{-\frac{3}{2}}\left(\frac{\partial \epsilon}{\partial c}\right)^2 + \epsilon^{-\frac{1}{2}}\frac{\partial^2 \epsilon}{\partial c^2}\right] \\ \frac{\partial E}{\partial c} &= 16c - 8a; & \frac{\partial^2 E}{\partial c^2} &= 16. \end{aligned}$$

In order to evaluate the expressions in (6.17) we must find

$$
\begin{aligned}
&\frac{\partial\beta}{\partial c} = -8c - bE^{-\frac{1}{2}}\frac{\partial E}{\partial c}; \qquad \frac{\partial^2\beta}{\partial c^2} = -8 - bE^{-\frac{3}{2}}\left[-\frac{1}{2}\left(\frac{\partial E}{\partial c}\right)^2 + E\frac{\partial^2 E;}{\partial c^2;}\right]\\
(6.18)\quad &\frac{\partial\gamma}{\partial c} = -8c + bE^{-\frac{1}{2}}\frac{\partial E}{\partial c}; \qquad \frac{\partial^2\gamma}{\partial c^2} = -8 + bE^{-\frac{3}{2}}\left[-\frac{1}{2}\left(\frac{\partial E}{\partial c}\right)^2 + E\frac{\partial^2 E}{\partial c^2}\right]\\
&\frac{\partial\epsilon}{\partial c} = 2(a + c); \qquad \frac{\partial^2\epsilon}{\partial c^2} = 2.
\end{aligned}
$$

If we now set $c = 0$, we obtain

$$
\begin{aligned}
&P = \tfrac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})\\
(6.19)\quad &\frac{\partial P}{\partial c} = \frac{a - (a^2 - b^2)^{\frac{1}{2}}}{2(a^2 - b^2)^{\frac{1}{2}}}\\
&\frac{\partial^2 P}{\partial c^2} = \frac{2a^4 - 4a^2b^2 + b^4 + (-2a^3 + 2ab^2)(a^2 - b^2)^{\frac{1}{2}}}{2b^2(a^2 - b^2)^{\frac{3}{2}}}.
\end{aligned}
$$

We may now substitute these values in (6.16) and then substitute the resulting expressions in (6.11). The remaining values that are required for (6.11) are easily computed since they may be obtained from φ with $c = 0$, i.e.

$$(6.20)\qquad {}_0\varphi_2(\theta_0, \theta_1, 0) = [\tfrac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}}]^{-\frac{1}{2}n}.$$

The result of these substitutions gives

$$(6.21)\qquad \frac{-n^2(n+3)P^{-\frac{1}{2}n-2}}{8(a^2 - b^2)^{\frac{1}{2}}},$$

in which we set $d = \frac{1}{2}(a - b)$ and $e = \frac{1}{2}(a + b)$ and integrate with respect to d and e. We obtain

$$(6.22)\qquad \frac{-n^2}{2(n+2)}[\tfrac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1},$$

and if we set $b = 0$ and integrate with respect to a, setting $a = 1$, $(\theta_0 = 0)$, we finally have

$$(6.23)\qquad E({}_0\lambda_2^{2/n}) = \frac{n}{n+2}.$$

We shall now obtain the first moment of λ_2 without the restriction that the mean equal zero. For this purpose let us consider

$$(6.24)\qquad \varphi_2(\theta_0, \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty}\cdots\int e^{-(1/2\sigma^2)[a_{00}(1-\theta_0) - a_{01}\theta_1 - a_{02}\theta_2]}\,\Pi\, dx_\alpha.$$

Here $a_1 = 1 - \theta_0 + m$, $a_2 = a_n = -\frac{1}{2}\theta_1 + m$, $a_3 = a_{n-1} = -\frac{1}{2}\theta_2 + m$, a_4 to $a_{n-2} = m$ where $m = (\theta_0 + \theta_1 + \theta_2)/n$. Expanding the determinant as in (6.12) we find

$$
\begin{aligned}
&\varphi_1^{-2}(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^{n}\sum_{i=1}^{n} a_i\omega_k^{i-1}\\
(6.25)\quad &\qquad = \prod_{k=1}^{n}\left(a_1 + 2a_2\cos\frac{2\pi k}{n} + 2a_3\cos\frac{4\pi k}{n} + \sum_{i=4}^{n-2} a_i\omega_k^{i-1}\right).
\end{aligned}
$$

Now

$$\sum_{i=4}^{n-2} a_i \omega_k^{i-1} = m \sum_{i=4}^{n-2} \omega_k^{i-1} \tag{6.26}$$

$$= \begin{cases} -m(1 + \omega^1 + \omega^{-1} + \omega^2 + \omega^{-2}), & k \neq n, \\ m(n - 5), & k = n, \end{cases}$$

so that

$$\varphi_2^{-2}(\theta_0, \theta_1, \theta_2) = \prod_{k=1}^{n-1} \left(a_1 + 2a_2 \cos \frac{2\pi k}{n} + 2a_3 \cos \frac{4\pi k}{n} \right). \tag{6.27}$$

We have obtained here a product which is the same as that in (6.12) except that the last factor is missing. The approximation corresponding to (6.14) will then be

$$\varphi_2(\theta_0, \theta_1, \theta_2) \cong \frac{[A + B + C + D]^{-\frac{1}{2}n}}{(a + b + c)^{-\frac{1}{2}}}, \tag{6.28}$$

since we may take the approximation for the product from 1 to n and divide by the last factor, $(a + b + c)$. The procedure for finding the first moment for λ_2 (mean $= a$) is exactly the same as that outlined for finding the first moment of ${}_0\lambda_2$ (mean $= 0$). We obtain

$$E(\lambda_2^{2/n}) = \frac{n - 1}{n + 1}. \tag{6.29}$$

Case III. If we set $r = 2$, $m = 1$ in $\lambda_{r,m}$ we have, if we take $l_1 = 1$, $l_2 = 2$

$$\lambda_{2,1}^{2/n} = \frac{\begin{vmatrix} a_{00} & a_{01} & a_{02} \\ a_{01} & a_{00} & a_{01} \\ a_{02} & a_{01} & a_{00} \end{vmatrix} a_{00}}{\begin{vmatrix} a_{00} & a_{01} \\ a_{01} & a_{00} \end{vmatrix}^2}. \tag{6.30}$$

To find the moments of $\lambda_{2,1}$ let us consider the following distribution,

$$dP_n = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{\alpha=1}^{n} [x_\alpha - \bar{x} - \beta(x_{\alpha+1} - \bar{x})]^2} \Pi\, dx_\alpha, \tag{6.31}$$

in which β represents the population value of the serial correlation coefficient. The moment generating function for the joint distribution of $a_{00}/2\sigma^2$, $a_{01}/2\sigma^2$ and $a_{02}/2\sigma^2$ will be

$$\begin{aligned} \varphi_{2,1}(\theta_0, \theta_1, \theta_2) &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int \exp\left(\frac{-1}{2\sigma^2} \{\Sigma[(x_\alpha - \bar{x}) - \beta(x_{\alpha+1} - \bar{x})]^2 \right. \\ &\qquad \left. - a_{00}\theta_0 - a_{01}\theta_1 - a_{02}\theta_2\}\right) \Pi\, dx_\alpha \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \int_{-\infty}^{\infty} \cdots \int \exp\left(-\frac{1}{2\sigma^2} [a_{00}(1 + \beta^2 - \theta_0) \right. \\ &\qquad \left. + a_{01}(-2\beta - \theta_1) + a_{02}(-\theta_2)]\right) \Pi\, dx_\alpha. \end{aligned} \tag{6.32}$$

This function is very similar to (6.24). The approximation to $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$ here will be exactly the same as that obtained in (6.28) for $\varphi_2(\theta_0, \theta_1, \theta_2)$ except that here $a = 1 + \beta^2 - \theta_0$, $b = -2\beta - \theta_1$, $c = -\theta_2$. For the case where the mean is zero, we find the approximation (6.14) in which a, b, and c have the above values.

We may obtain the first moment of $\lambda_{2,1}$ by operating on the function $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$ proceeding as follows: (a) compute (6.11) as before and set $\theta_2 = 0$, (b) integrate from $-\infty$ to $\theta_0 + \theta_1$ with respect to $\theta_0 + \theta_1 = \zeta$ from $-\infty$ to $\theta_0 - \theta_1$ with respect to $\theta_0 - \theta_1 = \xi$ (at this point we will have the first moment of the third order determinant divided by the second order determinant), (c) differentiate with respect to θ_0, (d) repeat step (b), and set θ_0 and $\theta_1 = 0$.

The first two steps for obtaining the first moment of ${}_0\lambda_{2,1}$ (mean $= 0$) were performed for the first moment of ${}_0\lambda_2$ so that we may perform step (c) on (6.22). We obtain

$$\frac{n^2[\frac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1}}{4(a^2 - b^2)^{\frac{1}{2}}}, \tag{6.33}$$

and finally by step (d) we have

$$E({}_0\lambda_{2,1}^{2/n}) = \frac{n}{n+1}\,[\tfrac{1}{2}(a + (a^2 - b^2)^{\frac{1}{2}})]^{-\frac{1}{2}n-1}, \tag{6.34}$$

in which $a = 1 + \beta^2$ and $b = -2\beta$ since θ_0 and θ_1 have been set equal to zero. Substitution of these values in (6.34) shows that it is independent of β, and we find

$$E({}_0\lambda_{2,1}^{2/n}) = \frac{n}{n+1}. \tag{6.35}$$

Using $\varphi_{2,1}(\theta_0, \theta_1, \theta_2)$, the generating function for $\lambda_{2,1}$ (mean $= a$), we find

$$E(\lambda_{2,1}^{2/n}) = \frac{n-1}{n}. \tag{6.36}$$

The procedure for obtaining the second moments of the above criteria consists essentially of performing twice the operations prescribed for obtaining the first moments. The details given in connection with the first moment are sufficient to indicate the procedure. The details for the second moments are too complicated algebraically to list here. Table I indicates the second moments obtained as well as other moments obtained in the earlier parts of the paper.

7. Serial correlation in several variables. Given a sample of n observations on each of k variables $x_{i\alpha}$, $i = 1, \cdots, k$, we shall assume they are distributed as follows:

$$dP_n = \frac{A^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}nk}}\, e^{-\frac{1}{2}\Sigma A_{ij}(x_{i\alpha} - a_i - b_i x_{i,\alpha+k})(x_{j\alpha} - a_j - b_j x_{j,\alpha+k})}\, \Pi\, dx_{i\alpha}. \tag{7.1}$$

We wish to test the hypothesis H_{kn} that there is no serial correlation, i.e., that $b_i = 0$, $i = 1, \cdots, k$. For this purpose let us define the space Ω and ω as follows:

$$\begin{cases} \Omega: & \| A_{ij} \| \text{ pos. def. } -\infty < a_i,\ b_i < \infty \\ \omega: & \| A_{ij} \| \text{ pos. def. } -\infty < a_i < \infty;\quad b_i = 0. \end{cases} \tag{7.2}$$

TABLE I

x	formula no.	$E(x)$	$E(x^2)$	σ^2
$\hat{b}_0$	(2.11)	0	$\frac{1}{n+2}$	$\frac{1}{n+2}$
$\hat{b}$	(2.6)	$\frac{-1}{n-1}$	$\frac{1}{n+1}$	$\frac{n(n-3)}{(n-1)^2(n+1)}$
${}_0\eta_1$	(5.3)	2	$\frac{4(n+3)}{n+2}$	$\frac{4}{n+2}$
η_1	(5.6)	$\frac{2n}{n-1}$	$\frac{4n(n+3)}{(n-1)(n+1)}$	$\frac{4n(n-3)}{(n-1)^2(n+1)}$
η_2	(5.8)	$\frac{3n+2}{n+1}$	$\frac{9n^2+23n+12}{(n+1)(n+2)}$	$\frac{2n^2+7n+4}{(n+1)^2(n+2)}$
${}_0\lambda_1$	(2.10)ff.	$\frac{n+1}{n+2}$	$\frac{(n+1)(n+3)}{(n+2)(n+4)}$	$\frac{2(n+1)}{(n+2)^2(n+4)}$
λ_1	(2.10)	$\frac{n}{n+1}$	$\frac{n(n+2)}{(n+1)(n+3)}$	$\frac{2n}{(n+1)^2(n+3)}$
${}_0\lambda_2$	(6.9)ff.	$\frac{n}{n+2}$	$\frac{n}{n+4}$	$\frac{4n}{(n+2)^2(n+4)}$
λ_2	(6.9)	$\frac{n-1}{n+1}$	$\frac{n-1}{n+3}$	$\frac{4(n-1)}{(n+1)^2(n+3)}$
${}_0\lambda_{2,1}$	(6.30)ff.	$\frac{n}{n+1}$	$\frac{n(n+2)}{(n+1)(n+3)}$	$\frac{2n}{(n+1)^2(n+3)}$
$\lambda_{2,1}$	(6.30)	$\frac{n-1}{n}$	$\frac{(n-1)(n+1)}{n(n+2)}$	$\frac{2(n-1)}{n^2(n+2)}$

The mean of $\hat{b}_0$ and $\hat{b}$ were also obtained by Anderson [8].

The development of the appropriate λ criterion for this case parallels very closely the development of the λ criteria in multiple regression analysis. The criterion obtained for testing the hypothesis H_{kn} is

$$\lambda_{kn}^{2/n} = \frac{\begin{vmatrix} a_{ij} & b_{ij} \\ b_{ij} & a_{ij} \end{vmatrix}}{|a_{ij}|^2}, \tag{7.3}$$

where

$$a_{ij} = \sum_\alpha (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j),$$

$$b_{ij} = \tfrac{1}{2}\left[\sum (x_{i\alpha} - \bar{x}_i)(x_{j,\alpha+l} - \bar{x}_j) + \sum (x_{i,\alpha+l} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)\right].$$

The probability theory for the λ criteria in (7.3) remains to be developed.

8. Summary. A problem in serial correlation which has received considerable attention is that of devising a statistic for indicating the presence of a relation between successive observations, i.e. a lack of independence of the order in which the observations were drawn. Von Neumann developed the distribution and moments of the ratio of the mean square successive difference to the variance. R. L. Anderson presented the distribution of a serial correlation coefficient which is the ratio $R = \Sigma x_\alpha x_{\alpha+l}/\Sigma x_\alpha^2$ ($l \geq 1$, subscripts reduced mod n).

The present investigation was undertaken with the object of developing the likelihood ratio functions for testing various hypotheses connected with serial correlation in one or more variables and determining the moments and in some cases the distributions of these likelihood ratios.

The variates are considered to be ordered by their subscripts $\alpha = 1, \cdots, n$. The introduction of $x_{n+1} = x_1$, $x_{n+2} = x_2$ etc. is made to obtain a symmetry which greatly simplifies the problem.

The likelihood ratio criteria were developed for testing the hypotheses

a) that x_α is independent of $x_{\alpha+l}$

b) that x_α is independent of $x_{\alpha+l_i}$, $i = 1, \cdots, r$;

c) that x_α is independent of some subset of the $x_{\alpha+l_i}$

d) that in the case of several variables $x_{i\alpha}$, $i = 1, \cdots, k$, $\alpha = 1, \cdots, n$ the $x_{i\alpha}$, $i = 1, \cdots, k$ are independent of the $x_{i,\alpha+h}$. These criteria are similar in form to those obtained in regression analysis.

The likelihood ratio criterion for testing the hypothesis *a*) turns out to be $\lambda = (1 - R^2)^{n/2}$ where R is the function given above. The moments of R are obtained and from these the moments of $\lambda^{2/n}$. These moments are found to agree with those of a Pearson Type I curve to $n/2$ moments. A simple transformation gives us the moments of a ratio differing from that used by von Neumann by the addition of the term $(x_n - x_1)^2$ to the numerator. A simplification of the moments is attained by this change. In fact, if we denote this altered statistic by η we find that $(\eta - 2)^2$ is distributed according to a Pearson Type I curve to $n/2$ moments.

The mean and variance were determined for the ratio of the sum of squares of the second successive differences to the first successive differences.

The mean and variance are obtained for the likelihood criteria for testing the hypothesis *b*) for $r = 2$, and for testing the hypothesis *c*) for $r = 2$ where $x_{\alpha+l_2}$ is the subset of $x_{\alpha+l_i}$; ($i = 1, 2$).

All the above moments were obtained under the assumption that the hypothesis to be tested was true. No results have been obtained thus far in cases *b*) and *c*) for a general r nor for hypothesis *d*).

The moments for the several cases above were obtained by the use of moment generating functions which, for the criteria used, took the form of the product of n terms. In the case *a*) it was shown that the product could be approximately represented by the nth power of a single expression which was equivalent for the purpose of obtaining the first n moments. A method was developed for making analogous approximations to the generating functions for cases *b*) and *c*) since

it was not found possible to obtain the moments from the products in their original form.

The author wishes to express his gratitude to Professor S. S. Wilks under whose helpful direction this paper was written.

REFERENCES

[1] J. von Neumann, R. H. Kent, H. R. Bellinson, and B. I. Hart, "The mean square successive difference," *Annals of Math. Stat.*, Vol. 12(1941) pp. 153–162.

[2] J. von Neumann, "Distribution of the ratio of the mean square difference to the variance," *Annals of Math. Stat.*, Vol. 12(1941) pp. 367–395.

[3] B. I. Hart and J. von Neumann, "Tabulation of the probabilities of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13(1942) pp. 207–214.

[4] J. von Neumann, "A further remark concerning the distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 13(1942) pp. 86–88.

[5] J. D. Williams, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Math. Stat.*, Vol. 12(1941) pp. 239–241.

[6] L. C. Young, "On randomness in ordered sequences," *Annals of Math. Stat.*, Vol. 12(1941) pp. 293–300.

[7] R. L. Anderson, "Serial correlation in the analysis of time series," unpublished thesis, Iowa State College, 1941.

[8] R. L. Anderson, "Distribution of the serial correlation coefficient," *Annals of Math. Stat.*, Vol. 13(1942) pp. 1–13.

[9] T. Koopmans, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13(1942) pp. 14–33.

[10] P. S. Laplace, *Traite de Mecanique Celeste*, Vol. 2, paragraph 21.

ON A STATISTICAL PROBLEM ARISING IN THE CLASSIFICATION OF AN INDIVIDUAL INTO ONE OF TWO GROUPS[1]

By Abraham Wald

Columbia University

1. Introduction. In social, economic and industrial problems we are often confronted with the task of classifying an individual into one of two groups on the basis of a number of test scores. For example, in the case of personnel selection the acceptance or rejection of an applicant is frequently based on a number of test scores obtained by the applicant. A similar situation arises in connection with college entrance examinations. Again, on the basis of a number of test scores, the admission or rejection of a student has to be decided. In all such problems it is assumed that there are two populations, say π_1 and π_2, one representing the population of individuals fit, and the other the population of individuals unfit for the purpose under consideration. The problem is that of classifying an individual into one of the populations π_1 and π_2 on the basis of his test scores. Often, some statistical data from past experience are available which can be utilized in making the classification. Suppose that from past experience we have the test scores of N_1 individuals who *are known* to belong to population π_1, and also the test scores of N_2 individuals who *are known* to belong to population π_2. These data will be utilized in classifying a new individual on the basis of his test scores.

In this paper we shall deal with the statistical problem of classifying an individual into one of the populations π_1 and π_2 on the basis of his test scores and on the basis of past experience, given in the form of two samples, one drawn from π_1 and the other from π_2. In the next section we give a precise formulation of the statistical problem and state the assumptions we make about the populations π_1 and π_2.

2. Statement of the problem. We consider two sets of p variates $(x_1, \cdots, x_p)$ and $(y_1, \cdots, y_p)$. It is assumed that each of the sets $(x_1, \cdots, x_p)$ and $(y_1, \cdots, y_p)$ has a p-variate normal distribution and the two sets are independent of each other. It is furthermore assumed that the covariance matrix of the variates $x_1, \cdots, x_p$ is equal to the covariance matrix of the variates $y_1, \cdots, y_p$, i.e. $\sigma_{x_i x_j} = \sigma_{y_i y_j}$ $(i, j = 1, \cdots, p)$. We will denote this common covariance by σ_{ij}. Let us denote the mean value of x_i by μ_i and the mean value of y_i by ν_i. Furthermore we will denote the normal population with mean values $\mu_1, \cdots, \mu_p$ and covariance matrix $\| \sigma_{ij} \|$ by π_1, and the normal population with mean values $\nu_1, \cdots, \nu_p$ and covariance matrix $\| \sigma_{ij} \|$ by π_2.

A sample of size N_1 is drawn from the population π_1 and a sample of size N_2 is

[1] The author wishes to thank Dr. Irving Lorge, Columbia University, for calling his attention to this problem.

drawn from the population π_2. Denote by $x_{i\alpha}$ the α-th observation on x_i ($i = 1, \cdots, p; \alpha = 1, \cdots, N_1$) and $y_{i\beta}$ the β-th observation on y_i ($i = 1, \cdots, p$; $\beta = 1, \cdots, N_2$). Let z_i ($i = 1, \cdots, p$) be a single observation on the i-th variate drawn from a p-variate population π, where it is known a priori that π is either identical with π_1 or with π_2. The set $(z_1, \cdots, z_p)$ is assumed to be distributed independently of $(x_1, \cdots, x_p)$ and $(y_1, \cdots, y_p)$.

We will deal here with the following statistical problem: On the basis of the observations $x_{i\alpha}, y_{i\beta}, z_i$ ($i = 1, \cdots, p$; $\alpha = 1, \cdots, N_1$; $\beta = 1, \cdots, N_2$) we test the hypothesis H_1 that the population π, from which the set $(z_1, \cdots, z_p)$ has been drawn, is equal to π_1. The parameters $\mu_1, \cdots, \mu_p, \nu_1, \cdots, \nu_p$ and $\| \sigma_{ij} \|$ are assumed to be unknown.

3. The statistic to be used for testing the hypothesis H_1. In this problem there exists only a single alternative hypothesis to the O-hypothesis H_1 to be tested, i.e. the hypothesis H_2 that π is equal to π_2. If the parameters $\mu_1, \cdots, \mu_p, \nu_1, \cdots, \nu_p$ and $\| \sigma_{ij} \|$ were known we could easily find (on the basis of a lemma by Neyman and Pearson) the critical region which is most powerful with respect to the alternative H_2. Let us assume for the moment that the parameters $\mu_1, \cdots, \mu_p, \nu_1, \cdots, \nu_p$ and $\| \sigma_{ij} \|$ are known and let us compute the critical region for testing H_1 which is most powerful with respect to the alternative H_2. According to a lemma by Neyman and Pearson[2] this critical region is given by the inequality

$$\frac{p_2(z_1, \cdots, z_p)}{p_1(z_1, \cdots, z_p)} \geq k, \tag{1}$$

where $p_1(z_1, \cdots, z_p)$ denotes the joint probability density function of $z_1, \cdots, z_p$ under the hypothesis H_1, $p_2(z_1, \cdots, z_p)$ denotes the joint probability density function of $(z_1, \cdots, z_p)$ under the hypothesis H_2, and k is a constant determined so that the critical region should have the required size.

Denote the determinant value $| \sigma_{ij} |$ of the matrix $\| \sigma_{ij} \|$ by σ^2. Then

$$p_1(z_1, \cdots, z_p) = \frac{1}{(2\pi)^{p/2}\sigma} e^{-\frac{1}{2}\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma^{ij}(z_i-\mu_i)(z_j-\mu_j)}, \tag{2}$$

and

$$p_2(z_1, \cdots, z_p) = \frac{1}{(2\pi)^{p/2}\sigma} e^{-\frac{1}{2}\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma^{ij}(z_i-\nu_i)(z_j-\nu_j)}, \tag{3}$$

where the matrix $\| \sigma^{ij} \|$ denotes the inverse matrix of the matrix $\| \sigma_{ij} \|$. Taking logarithms of both sizes of the inequality (1), we obtain the inequality

$$-\tfrac{1}{2}\left\{\sum_j \sum_i \sigma^{ij}[(z_i - \nu_i)(z_j - \nu_j) - (z_i - \mu_i)(z_j - \mu_j)]\right\} \geq \log k. \tag{4}$$

[2] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses," *Stat. Res. Mem.*, Vol. 1, London, 1936.

Multiplying both sides of (4) by 2, we have

$$\sum_j \sum_i \sigma^{ij}[(z_i - \mu_i)(z_j - \mu_j) - (z_i - \nu_i)(z_j - \nu_j)] \geq 2 \log k. \tag{5}$$

The critical region (5) is most powerful with respect to the alternative H_2, but it cannot be used for our purposes since the parameters $\mu_1, \cdots, \mu_p, \nu_1, \cdots, \nu_p$ and $\| \sigma_{ij} \|$ are unknown. The optimum estimate of σ_{ij} on the basis of the observations $x_{i\alpha}$ and $y_{i\beta}$ is given by the sample covariance

$$s_{ij} = \frac{\sum_{\alpha=1}^{N_1} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j) + \sum_{\beta=1}^{N_2} (y_{i\beta} - \bar{y}_i)(y_{j\beta} - \bar{y}_j)}{N_1 + N_2 - 2} \tag{6}$$

where $\bar{x}_i = \dfrac{\sum_\alpha x_{i\alpha}}{N_1}$ and $\bar{y}_i = \dfrac{\sum_\beta y_{i\beta}}{N_2}$. The optimum estimates of μ_i and ν_i are given by $\bar{x}_i$ and $\bar{y}_i$ respectively $(i = 1, \cdots, p)$. Hence for testing H_1 it seems reasonable to use the statistic R which we obtain from the left hand side of (5) by substituting the optimum estimates for the unknown parameters. Thus R is given by

$$R = \sum_j \sum_i s^{ij}[(z_i - \bar{x}_i)(z_j - \bar{x}_j) - (z_i - \bar{y}_i)(z_j - \bar{y}_j)], \tag{7}$$

where $\| s^{ij} \| = \| s_{ij} \|^{-1}$. The critical region for testing H_1 is given by the inequality

$$R \geq C, \tag{8}$$

where C is a constant determined in such a way that the critical region should have the required size. It is interesting to notice that R is proportional to the difference $T_1^2 - T_2^2$ where T_i $(i = 1, 2)$ denotes the generalized Student's ratio[3] for testing the hypothesis that the set $(z_1, \cdots, z_p)$ is drawn from the population π_i. In our case the statistic T_1 cannot be used for testing H_1, since T_1 is appropriate for this purpose if the class of alternative hypotheses contains *all* p-variate normal populations having the same covariance matrix as π_1. In our case the class of alternatives consists merely of a single alternative, namely, the alternative π_2.

For the sake of certain simplifications we shall propose the use of a statistic U which differs slightly from the statistic R. In order to obtain U, we consider the inequality (5). Since $\sigma^{ij} = \sigma^{ji}$ this inequality can be reduced to

$$\sum_j \sum_i \sigma^{ij} z_i(\nu_j - \mu_j) \geq k', \tag{9}$$

where k' denotes a certain constant. The statistic U is obtained from the left hand side of (9) by substituting the optimum estimates for the unknown para-

[3] See. in this connection H. Hotelling, "The generalization of Student's ratio," *Annals of Math. Stat.*, Vol. 2, and R. C. Bose and S. N. Roy, "The exact distribution of the Studentized D^2 statistic," *Sankhya*, Vol. 3.

meters. Thus

$$U = \Sigma\Sigma s^{ij} z_i(\bar{y}_j - \bar{x}_j), \tag{10}$$

and the critical region is given by the inequality

$$U \geq d, \tag{11}$$

where the constant d is chosen so that the critical region should have the required size. The statistic U differs from R merely by a term which does not depend on the quantities $z_1, \cdots, z_p$. If N_1 and N_2 are large the difference $U - R$ is practically constant and therefore the critical regions (8) and (11) are identical. The use of U seems to be as justifiable as that of R and because of certain simplifications we propose the use of the critical region (11).

The statistic U is closely connected with the so called discriminant function[4] introduced by R. A. Fisher for discriminating between the two populations π_1 and π_2. The discriminant function D is given by

$$D = b_1 d_1 + b_2 d_2 + \cdots + b_p d_p \tag{12}$$

where $d_i = \bar{y}_i - \bar{x}_i$ and the coefficient b_i is proportional to $\sum_{j=1}^{p} s^{ij} d_j$. The coefficients $b_1, \cdots, b_p$ are called the coefficients of the discriminant function. We see that U is proportional to the statistic $\sum_{i=1}^{p} b_i z_i$ which is obtained from the right hand side of (12) by substituting z_i for d_i.

4. Solution of the problem when N_1 and N_2 are large. Denote by $F(U, N_1, N_2 \mid \pi_i)$ the cumulative probability distribution of U under the hypothesis that the set $(z_1, \cdots, z_p)$ has been drawn from the population π_i $(i = 1, 2)$. If N_1 and N_2 approach infinity the distribution $F(U, N_1, N_2 \mid \pi_i)$ converges to a normal distribution, since the variates s_{ij}, $\bar{x}_i$ and $\bar{y}_i$ converge stochastically to the constants σ_{ij}, μ_i and ν_i respectively $(i, j = 1, \cdots, p)$. Let us denote $\lim_{N_1 = N_2 = \infty} F(U, N_1, N_2 \mid \pi_i)$ by $\Phi(U \mid \pi_i)$ $(i = 1, 2)$. Furthermore denote by α_i the mean value, and by σ_i the standard deviation of the distribution $\Phi(U \mid \pi_i)$ $(i = 1, 2)$. It is obvious that $\sigma_1 = \sigma_2 = \sigma$ (say). It is easy to verify that the variates

$$\bar{\alpha}_1 = \Sigma\Sigma s^{ij} \bar{x}_i(\bar{y}_j - \bar{x}_j), \tag{13}$$

$$\bar{\alpha}_2 = \Sigma\Sigma s^{ij} \bar{y}_i(\bar{y}_j - \bar{x}_j), \tag{14}$$

$$\begin{aligned} \bar{\sigma}^2 &= \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} s^{ik} s^{jl} (\bar{y}_k - \bar{x}_k)(\bar{y}_l - \bar{x}_l) s_{ij} \\ &= \sum_{k=1}^{p}\sum_{l=1}^{p} s^{kl} (\bar{y}_k - \bar{x}_k)(\bar{y}_l - \bar{x}_l), \end{aligned} \tag{15}$$

converge stochastically to the constants α_1, α_2 and σ^2 respectively.

[4] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, 1938.

Hence for large values of N_1 and N_2 we can assume that U is normally distributed with mean value $\bar{\alpha}_i$ and standard deviation $\bar{\sigma}$ if the hypothesis H_i ($i = 1, 2$) is true. Thus the critical region for testing H_1 is given by the inequality

$$U \geq \bar{\alpha}_1 + \lambda\bar{\sigma}, \tag{16}$$

where the constant λ is chosen in such a way that $\frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-t^2/2}\, dt$ is equal to the required size of the critical region.

Finally, some remarks about the proper choice of the size of the critical region may be of interest. Two kinds of error may be committed. H_1 may be rejected when it is true, and H_1 may be accepted when H_2 is true. Suppose that W_1 and W_2 are two positive numbers expressing the importance of an error of the first kind and an error of the second kind respectively. If the purpose of the statistical investigation is given it will usually be possible to determine the values of W_1 and W_2. We shall deal here with the question of determining the size of the critical region as a function of the weights W_1 and W_2. Denote by P_i the probability that (16) holds under the assumption that H_i is true ($i = 1, 2$). Then P_1 is the size of the critical region (also the probability of an error of the first kind), and $1 - P_2$ is the probability of an error of the second kind. Both probabilities P_1 and P_2 are functions of λ and are given by the following expressions:

$$P_1 = \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-t^2/2}\, dt, \tag{17}$$

and

$$P_2 = \frac{1}{\sqrt{2\pi}} \int_{((\bar{\alpha}_1 - \bar{\alpha}_2)/\bar{\sigma})+\lambda}^\infty e^{-t^2/2}\, dt. \tag{18}$$

From (13) and (14) we obtain

$$\bar{\alpha}_2 - \bar{\alpha}_1 = \sum_j \sum_i s^{ij}(\bar{y}_i - \bar{x}_i)(\bar{y}_j - \bar{x}_j). \tag{19}$$

Since the right hand side of (19) is positive definite, we have $\bar{\alpha}_2 > \bar{\alpha}_1$. Hence because of (17) and (18) we also have $P_2 > P_1$. By the risk of committing a certain error we understand the probability of that error multiplied by its weight. Hence the risk of committing an error of the first kind is given by W_1P_1, and the risk of committing an error of the second kind is given by $W_2(1 - P_2)$. It seems reasonable to choose the value of λ so that the two risks become equal to each other, i.e. such that

$$W_1P_1 = W_2(1 - P_2). \tag{20}$$

Hence using (17) and (18) we obtain the following equation in λ

$$W_1 \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-t^2/2}\, dt - W_2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{((\bar{\alpha}_1 - \bar{\alpha}_2)/\bar{\sigma})+\lambda} e^{-t^2/2}\, dt = 0. \tag{21}$$

Using a table of the normal distribution, the value of λ which satisfies the equation (21) can easily be found. For $W_1 = W_2$ the solution of (21) is given by

$$\lambda = \frac{\bar{\alpha}_2 - \bar{\alpha}_1}{2\bar{\sigma}},$$

and the critical region is given by the inequality

$$U \geq \bar{\alpha}_1 + \lambda\bar{\sigma} = \bar{\alpha}_1 + \frac{\bar{\alpha}_2 - \bar{\alpha}_1}{2} = \frac{\bar{\alpha}_1 + \bar{\alpha}_2}{2}.$$

5. Some results concerning the exact sampling distribution of the statistic U. If N_1 and N_2 are not large the solution given in section 4 cannot be used and it is necessary to derive the exact sampling distribution of U. Let

$$(\bar{y}_i - \bar{x}_i)\sqrt{\frac{N_1 N_2}{N_1 + N_2}} = z'_i \qquad (i = 1, \cdots, p). \tag{22}$$

Then

$$U = \sqrt{\frac{N_1 + N_2}{N_1 N_2}} \sum_i \sum_j s^{ij} z_i z'_j \tag{23}$$

where the variates $z'_1, \cdots, z'_p$ are distributed independently of the set $(z_1, \cdots, z_p)$, the mean value of z'_i is equal to $(\nu_i - \mu_i)\sqrt{\frac{N_1 N_2}{N_1 + N_2}}$ and the covariance between z'_i and z'_j is equal to σ_{ij}. It is known that the set of covariances s_{ij} is distributed independently of the set $(z_1, \cdots, z_p, z'_1, \cdots, z'_p)$ and therefore the distribution of U remains unchanged if instead of (6) we have

$$s_{ij} = \frac{\sum_{\alpha=1}^{n} t_{i\alpha}^2}{n} \qquad (n = N_1 + N_2 - 2), \tag{24}$$

where the variates $t_{i\alpha}$ are distributed independently of the set $(z_1, \cdots, z_p, z'_1, \cdots, z'_p)$, have a joint normal distribution with mean values zero, $\sigma_{t_{i\alpha}t_{j\alpha}} = \sigma_{ij}$ and $\sigma_{t_{i\alpha}t_{j\beta}} = 0$ if $\alpha \neq \beta$. It is necessary to derive the distribution of U under both hypotheses H_1 and H_2. In both cases the mean values of $z_1, \cdots, z_p, z'_1, \cdots, z'_p$ are not zero. Instead of U we will consider the statistic

$$U' = \sum_{i=1}^{p} \sum_{j=1}^{p} s^{ij} z_i z'_j$$

which differs from U only in the proportionality factor $\sqrt{\frac{N_1 + N_2}{N_1 N_2}}$. The distributions of U' under the hypotheses H_1 and H_2 are contained as special cases in the distribution of the statistic

$$V = \sum_j \sum_i s^{ij} t_{i,n+1} t_{j,n+2}, \tag{25}$$

where s_{ij} is given by (24) and the joint distribution of the variates $t_{i\beta}$ $(i = 1, \cdots, p; \beta = 1, \cdots, n + 2)$ is given by

$$\frac{1}{(2\pi)^{p(n+2)/2}\sigma^{n+2}} e^{-\frac{1}{2}\sum_{j=1}^{p}\sum_{i=1}^{p}\sigma^{ij}\left[\sum_{\alpha=1}^{n} t_{i\alpha}t_{j\alpha}+(t_{i,n+1}-\xi_i)(t_{j,n+1}-\xi_j)+(t_{i,n+2}-\eta_i)(t_{j,n+2}-\eta_j)\right]} \tag{26}$$

$$\times \prod_{\beta=1}^{n+2}\sum_{i=1}^{p} dt_{i\beta}.$$

The quantities $\xi_1, \cdots, \xi_p, \eta_1, \cdots, \eta_p$ are constants and σ^2 denotes the determinant value of the matrix $\| \sigma_{ij} \|$.

We will deal here with the distribution of the statistic V given in (25) under the assumption that the joint distribution of the variates $t_{i\beta}$ $(i = 1, \cdots, p; \beta = 1, \cdots, n + 2)$ is given by (26).

In order to derive the distribution of V we shall have to prove several lemmas.

LEMMA 1. *Let $\| \lambda_{ij} \|$ $(i, j = 1, \cdots, p)$ be an arbitrary non-singular matrix, and let*

$$t'_{i\beta} = \sum_{j=1}^{p} \lambda_{ij} t_{j\beta} \qquad (i = 1, \cdots, p; \beta = 1, \cdots, n + 2).$$

Let furthermore s'_{ij} be given by

$$s'_{ij} = \frac{\sum_{\alpha=1}^{n} t'_{i\alpha} t'_{j\alpha}}{n}.$$

Then $\sum_j \sum_i s^{ij} t_{i,n+1} t_{j,n+2} = \sum_j \sum_i s'^{ij} t'_{i,n+1} t'_{j,n+2}$, i.e. the statistic V is invariant under non-singular linear transformations.

PROOF. We obviously have

$$t'_{i,n+1} t'_{j,n+2} = \sum_{k=1}^{p} \sum_{l=1}^{p} \lambda_{ik} \lambda_{jl} t_{k,n+1} t_{l,n+2}. \tag{27}$$

Furthermore we have

$$s'_{ij} = \sum_{k=1}^{p} \sum_{l=1}^{p} \lambda_{ik} \lambda_{jl} s_{k.}. \tag{28}$$

Hence

$$\| s'_{ij} \| = \| \lambda_{ij} \| \, \| s_{ij} \| \, \| \bar{\lambda}_{ij} \| \tag{29}$$

where $\bar{\lambda}_{ij} = \lambda_{ji}$.

From (29) we obtain

$$\| s'^{ij} \| = \| \bar{\lambda}^{ij} \| \, \| s^{ij} \| \, \| \lambda^{ij} \|, \tag{30}$$

and therefore

$$s'^{ij} = \sum_{k=1}^{p} \sum_{l=1}^{p} \lambda^{ki} \lambda^{lj} s^{kl}. \tag{31}$$

Hence from (27) and (31) we obtain

$$(32)\qquad \sum_j \sum_i s'^{ij} t'_{i,n+1} t'_{j,n+2} = \sum_j \sum_i \sum_k \sum_l \sum_u \sum_v \lambda^{ki} \lambda^{lj} s^{kl} \lambda_{iu} \lambda_{jv} t_{u,n+1} t_{v,n+2}.$$

The coefficient of $t_{u,n+1}t_{v,n+2}$ on the right hand side of (32) is given by

$$(33)\qquad \sum_j \sum_i \sum_k \sum_l \lambda^{ki} \lambda^{lj} s^{kl} \lambda_{iu} \lambda_{jv} = \sum_k \sum_l \{(\sum_i \lambda^{ki} \lambda_{iu})(\sum_j \lambda^{lj} \lambda_{jv}) s^{kl}\} = s^{uv}.$$

Lemma 1 follows from (32) and (33).

LEMMA 2. *The distribution of V remains unchanged if we assume that the covariance matrix $\| \sigma_{ij} \|$ is equal to the unit matrix, i.e. the joint distribution of the variates $t_{i\beta}$ $(i = 1, \cdots, p; \beta = 1, \cdots, n + 2)$ is given by*

$$(34)\qquad \frac{1}{(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2}\left[\sum_{i=1}^{p} \sum_{\alpha=1}^{n} t_{i\alpha}^2 + \sum_i (t_{i,n+1}-\rho_i)^2 + \sum_i (t_{i,n+2}-\zeta_i)^2\right]},$$

where the constants ρ_i and ζ_i are functions of the constants $\xi_1, \cdots, \xi_p, \eta_1, \cdots, \eta_p$ and of the σ_{ij}.

Lemma 2 is an immediate consequence of Lemma 1. Hence we have to derive the distribution of V under the assumption that the variates $t_{i\beta}$ have the joint distribution given in (34).

Let R_i $(i = 1, \cdots, p)$ be the point of the $n + 2$ dimensional Cartesian space with the coordinates $t_{i1}, \cdots, t_{i,n+2}$. Let $P = (u_1, \cdots, u_{n+2})$ and $Q = (v_1, \cdots, v_{n+2})$ be two arbitrary points such that $\sum_{\beta=1}^{n+2} u_\beta v_\beta = 0$ and $\sum u_\beta^2 = \sum v_\beta^2 = 1$. Denote by 0 the origin of the coordinate system and let $\bar{t}_{i,n+1}$ be the projection of the vector $0R_i$ on the vector $0P$. We have

$$(35)\qquad \bar{t}_{i,n+1} = \sum_{\beta=1}^{n+2} t_{i\beta} u_\beta \qquad (i = 1, \cdots, p).$$

Similarly, the projection $\bar{t}_{i,n+2}$ of the vector $0R_i$ on $0Q$ is given by

$$(36)\qquad \bar{t}_{i,n+2} = \sum_{\beta=1}^{n+2} t_{i\beta} v_\beta .$$

Let $\bar{R}_i$ $(i = 1, \cdots, p)$ be the projection of the point R_i on the n-dimensional hyperplane through 0 and perpendicular to the vectors $0P$ and $0Q$. Denote the coordinates of $\bar{R}_i$ by $r_{i1}, \cdots, r_{i,n+2}$ respectively and let $\bar{s}_{ij}$ be defined by

$$(37)\qquad \bar{s}_{ij} = \frac{\sum_{\beta=1}^{n+2} r_{i\beta} r_{j\beta}}{n}.$$

If we rotate the coordinate system so that the $(n + 1)$-axis coincides with $0P$ and the $(n + 2)$-axis coincides with $0Q$, and if $\bar{t}_{i1}, \cdots, \bar{t}_{i,n+2}$ denote the coordinates of R_i $(i = 1, \cdots, p)$ referred to the new system, then we have

$$(38)\qquad \bar{s}_{ij} = \frac{1}{n} \sum_{\beta=1}^{n+2} r_{i\beta} r_{j\beta} = \frac{1}{n} \sum_{\alpha=1}^{n} \bar{t}_{i\alpha} \bar{t}_{j\alpha}, \text{ and}$$

$$(39)\qquad \sum_{\beta=1}^{n+2} t_{i\beta} t_{j\beta} = \sum_{\beta=1}^{n+2} \bar{t}_{i\beta} \bar{t}_{j\beta} .$$

From (38) and (39) we obtain

$$\bar{s}_{ij} = \frac{\sum_{\beta=1}^{n+2} t_{i\beta} t_{j\beta} - \bar{t}_{i,n+1}\bar{t}_{j,n+1} - \bar{t}_{i,n+2}\bar{t}_{j,n+2}}{n}. \tag{40}$$

We will now prove

Lemma 3. *Let $\bar{V}$ be defined by*

$$\bar{V} = \sum_j \sum_i \bar{s}^{ij} \bar{t}_{i,n+1} \bar{t}_{j,n+2}, \tag{41}$$

where $\bar{t}_{i,n+1}$, $\bar{t}_{i,n+2}$ and $\bar{s}_{ij}$ are given by the formulas (35), (36) and (40) respectively. Let furthermore the joint probability distribution of the variates $t_{i\beta}$ ($i = 1, \cdots, p$; $\beta = 1, \cdots, n+2$) be given by

$$\frac{1}{(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2}\left[\sum_{i=1}^{p} \sum_{\beta=1}^{n+2} (t_{i\beta} - \rho_i u_\beta - \zeta_i v_\beta)^2\right]} \prod_i \prod_\beta dt_{i\beta}. \tag{42}$$

Then the distribution of $\bar{V}$ calculated under the assumption that the quantities $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ are constants and the joint probability distribution of the variates $t_{i\beta}$ is given by (42), is the same as the distribution of V calculated under the assumption that the joint probability distribution of the variates $t_{i\beta}$ is given by (34).

Proof. If we rotate the coordinate system so that the $(n+1)$-axis coincides with $0P$ and the $(n+2)$-axis coincides with $0Q$, and if $\bar{t}_{i1}, \cdots, \bar{t}_{i,n+2}$ denote the coordinates of R_i ($i = 1, \cdots, p$) in the new system, then $\bar{t}_{i,n+1}$ and $\bar{t}_{i,n+2}$ are given by the right hand sides of (35) and (36) respectively. Furthermore

$$\bar{s}_{ij} = \frac{\sum_{\alpha=1}^{n} \bar{t}_{i\alpha}\bar{t}_{j\alpha}}{n}.$$

Hence the distribution of $\bar{V}$ is certainly the same as that of V if the joint probability distribution of the variates $\bar{t}_{i\beta}$ ($i = 1, \cdots, p$; $\beta = 1, \cdots, n+2$) is given by the expression which we obtain from (34) by substituting $\bar{t}_{i\beta}$ for $t_{i\beta}$. Thus, in order to prove Lemma 3 we have merely to show that if the variates $\bar{t}_{i\beta}$ have the joint probability distribution (34), the variates $t_{i\beta}$ have the joint probability distribution (42). Since the variates $t_{i1}, \cdots, t_{i,n+2}$ are obtained by an orthogonal transformation of the variates $\bar{t}_{i1}, \cdots, \bar{t}_{i,n+2}$, it follows that the variates $t_{i\beta}$ ($i = 1, \cdots, p$; $\beta = 1, \cdots, n+2$) are independently and normally distributed with unit variances. We have

$$t_{i\beta} = \sum_{\gamma=1}^{n+2} \lambda_{\beta\gamma} \bar{t}_{i\gamma} \tag{43}$$

where $\lambda_{\beta\gamma}$ is equal to the cosine of the angle between the β-th axis of the original system and γ-th axis of the new system. Since

$$\lambda_{\beta,n+1} = u_\beta \quad \text{and} \quad \lambda_{\beta,n+2} = v_\beta,$$

and since $E(\bar{t}_{i\gamma}) = 0$ for $\gamma = 1, \cdots, n$, $E(\bar{t}_{i,n+1}) = \rho_i$ and $E(\bar{t}_{i,n+2}) = \zeta_i$, it follows from (43) that

$$E(t_{i\beta}) = \rho_i u_\beta + \zeta_i v_\beta . \tag{44}$$

Hence Lemma 3 is proved.

We will now prove

LEMMA 4. *Let P be a point with the coordinates $u_1, \cdots, u_{n+2}$ and Q a point with the coordinates $v_1, \cdots, v_{n+2}$ such that $\Sigma u_\beta v_\beta = 0$ and $\Sigma u_\beta^2 = \Sigma v_\beta^2 = 1$. Denote by L_p the flat space determined by the vectors $0R_1, \cdots, 0R_p$ ($R_i = (t_{i1}, \cdots, t_{i,n+2})$) and let $\bar{P}$ be the projection of P on L_p and $\bar{Q}$ the projection of Q on L_p. Denote furthermore by θ_1 the angle between the vectors $0P$ and $0\bar{P}$, by θ_1' the angle between $0P$ and $0\bar{Q}$, by θ_2 the angle between $0Q$ and $0\bar{Q}$, by θ_2' the angle between $0Q$ and $0\bar{P}$, and finally by θ_3 the angle between $0\bar{P}$ and $0\bar{Q}$. Then the statistic $\bar{V}$ defined in* (41) *is equal to*

$$\bar{V} = -\frac{\begin{vmatrix} 0 & a_1 & a_2 \\ b_1 & a_{11} & a_{12} \\ b_2 & a_{12} & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}}, \tag{45}$$

where

$$a_1 = \cos^2\theta_1; \quad a_2 = \cos\theta_1'\cos\theta_2; \quad b_1 = \cos\theta_1\cos\theta_2'; \quad b_2 = \cos^2\theta_2; \tag{46}$$

$$a_{11} = \frac{\cos^2\theta_1 - a_1^2 - b_1^2}{n}, \qquad a_{22} = \frac{\cos^2\theta_2 - a_2^2 - b_2^2}{n}$$
$$\text{and} \quad a_{12} = \frac{\cos\theta_1\cos\theta_2\cos\theta_3 - a_1a_2 - b_1b_2}{n}. \tag{47}$$

PROOF. If we rotate the coordinate system in such a way that the $(n+1)$-axis coincides with $0P$ and the $(n+2)$-axis coincides with $0Q$, and if $\bar{t}_{i1}, \cdots, \bar{t}_{i,n+2}$ are the coordinates of R_i in the new system, then

$$\bar{s}_{ij} = \frac{\sum_{\alpha=1}^{n} \bar{t}_{i\alpha}\bar{t}_{j\alpha}}{n}.$$

According to Lemma 1 the statistic V is invariant under linear transformations of the variables $t_{i\beta}$. Hence $\bar{V}$ is also invariant under linear transformations of the variables $\bar{t}_{i\beta}$. Thus the value of $\bar{V}$ remains unchanged if the points $R_1, \cdots, R_p$ are replaced by arbitrary points $R_1', \cdots, R_p'$ of L_p subject to the condition that the vectors $0R_1', \cdots, 0R_p'$ be linearly independent. Hence we may assume that the vectors $0R_3, \cdots, 0R_p$ are perpendicular to each other and lie in the intersection of L_p with the n-dimensional flat space which goes through 0 and is perpendicular to $0P$ and $0Q$. Furthermore we may assume that $R_1 = \bar{P}$ and $R_2 = \bar{Q}$. Then $0R_i$ is perpendicular to $0P$, $0Q$, $0R_1$ and $0R_2$ $(i = 3, \cdots, p)$.

The statistic $\bar{V}$ can obviously be written in the form:

$$\bar{V} = -\frac{\begin{vmatrix} 0 & \bar{l}_{1,n+1} & \cdots & \bar{l}_{p,n+1} \\ \bar{l}_{1,n+2} & \bar{s}_{11} & \cdots & \bar{s}_{1p} \\ \vdots & \vdots & & \vdots \\ \bar{l}_{p,n+2} & \bar{s}_{p1} & \cdots & \bar{s}_{pp} \end{vmatrix}}{\begin{vmatrix} \bar{s}_{11} & \cdots & \bar{s}_{1p} \\ \vdots & & \vdots \\ \bar{s}_{p1} & \cdots & \bar{s}_{pp} \end{vmatrix}}. \tag{48}$$

Because of our choice of the points $R_1, \cdots, R_p$, we have

$$\bar{l}_{i,n+1} = \bar{l}_{i,n+2} = 0 \qquad (i = 3, \cdots, p) \tag{49}$$

and

$$\sum_{\beta=1}^{n+2} \bar{l}_{i\beta}\bar{l}_{j\beta} = 0 \quad \text{if} \quad i \neq j \qquad (i = 3, \cdots, p; j = 1, \cdots, p). \tag{50}$$

From (49) and (50) it follows that $\bar{s}_{ij} = 0$ for $i \neq j$ except $\bar{s}_{12}$ which is not necessarily zero. Hence $\bar{V}$ reduces to the expression

$$\bar{V} = -\frac{\begin{vmatrix} 0 & \bar{l}_{1,n+1} & \bar{l}_{2,n+1} \\ \bar{l}_{1,n+2} & \bar{s}_{11} & \bar{s}_{12} \\ \bar{l}_{2,n+2} & \bar{s}_{12} & \bar{s}_{22} \end{vmatrix}}{\begin{vmatrix} \bar{s}_{11} & \bar{s}_{12} \\ \bar{s}_{12} & \bar{s}_{22} \end{vmatrix}}. \tag{51}$$

We obviously have $\bar{l}_{1,n+1} = a_1$, $\bar{l}_{2,n+1} = a_2$, $\bar{l}_{1,n+2} = b_1$ and $\bar{l}_{2,n+2} = b_2$.

For any two points A and B denote the length of the vector AB by $\overline{AB}$. Since $n\bar{s}_{11} + (\bar{l}_{1,n+1})^2 + (\bar{l}_{1,n+2})^2 = \overline{0P}^2$, $n\bar{s}_{22} + (\bar{l}_{2,n+1})^2 + (\bar{l}_{2,n+2})^2 = \overline{0Q}^2$ and $n\bar{s}_{12} + \bar{l}_{1,n+1}\bar{l}_{2,n+1} + \bar{l}_{1,n+2}\bar{l}_{2,n+2} = \overline{0P}\cdot\overline{0Q}\cdot\cos\theta_3$, we can easily verify that $\bar{s}_{11} = a_{11}$, $\bar{s}_{12} = a_{12}$ and $\bar{s}_{22} = a_{22}$. Hence Lemma 4 is proved.

The angles θ_1' and θ_2' can be expressed in terms of the angles θ_1, θ_2 and θ_3. In order to show this, let us rotate the coordinate system so that the first p coordinates lie in the flat space L_p defined in Lemma 4. Let $u_1', \cdots, u_{n+2}'$ be the coordinates of P and $v_1', \cdots, v_{n+2}'$ the coordinates of Q referred to the new axes. Then, since $\overline{0P} = \overline{0Q} = 1$, we have

$$\cos\theta_1 = \sqrt{u_1'^2 + \cdots + u_p'^2}; \qquad \cos\theta_1' = \frac{u_1'v_1' + \cdots + u_p'v_p'}{\sqrt{v_1'^2 + \cdots + v_p'^2}};$$

$$\cos\theta_2 = \sqrt{v_1'^2 + \cdots + v_p'^2}; \qquad \cos\theta_2' = \frac{u_1'v_1' + \cdots + u_p'v_p'}{\sqrt{u_1'^2 + \cdots + u_p'^2}};$$

and
$$\cos\theta_3 = \frac{u_1'v_1' + \cdots + u_p'v_p'}{\sqrt{u_1'^2 + \cdots + u_p'^2}\sqrt{v_1'^2 + \cdots + v_p'^2}}.$$

Hence

$$\cos\theta_1' = \cos\theta_1 \cos\theta_3 \quad \text{and} \quad \cos\theta_2' = \cos\theta_2 \cos\theta_3.$$

Introducing the notations

$$m_1 = \cos^2 \theta_1, \quad m_2 = \cos^2 \theta_2 \quad \text{and} \quad m_3 = \cos \theta_1 \cos \theta_2 \cos \theta_3,$$

we have

$$a_1 = m_1, \qquad a_2 = m_3, \qquad b_1 = m_3, \qquad b_2 = m_2;$$

$$\begin{cases} a_{11} = \dfrac{m_1 - m_1^2 - m_3^2}{n}, & a_{12} = \dfrac{m_3(1 - m_1 - m_2)}{n} \\ \text{and} & a_{22} = \dfrac{m_2 - m_2^2 - m_3^2}{n} \end{cases}.$$

Substituting the above values in (45) we obtain

$$\begin{aligned} \bar{V} &= -n \frac{m_3}{m_3^2 - 1 + m_1 + m_2 - m_1 m_2} \\ &= -n \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3}{\cos^2 \theta_1 \cos^2 \theta_2 \cos^2 \theta_3 - \sin^2 \theta_1 \sin^2 \theta_2}. \end{aligned}$$

Hence, Lemma 4 can be written as

LEMMA 4'. *Let P be a point with the coordinates $u_1, \cdots, u_{n+2}$ and Q a point with the coordinates $v_1, \cdots, v_{n+2}$. Denote by L_p the flat space determined by the vectors $0R_1, \cdots, 0R_p$ and let $\bar{P}$ be the projection of P on L_p and $\bar{Q}$ the projection of Q on L_p. Denote furthermore by θ_1 the angle between $0P$ and $0\bar{P}$, by θ_2 the angle between $0Q$ and $0\bar{Q}$ and by θ_3 the angle between $0\bar{P}$ and $0\bar{Q}$. Then the statistic $\bar{V}$ defined in* (41) *is equal to*

$$\bar{V} = -n \frac{\cos \theta_1 \cos \theta_2 \cos \theta_3}{\cos^2 \theta_1 \cos^2 \theta_2 \cos^2 \theta_3 - \sin^2 \theta_1 \sin^2 \theta_2}. \tag{45'}$$

If P is a point of the $(n + 1)$-axis and Q a point of the $(n + 2)$-axis, then $\bar{V}$ is identical with the statistic V given in (25). Hence we obtain the following

Geometric interpretation of the statistic V defined in (25). If θ_1 denotes the angle between the $(n + 1)$-axis and the flat space L_p determined by the vectors $0R_1, \cdots, 0R_p$, θ_2 the angle between the $(n + 2)$-axis and the flat space L_p, and if θ_3 denotes the angle between the projections of the last two coordinate axes on L_p, then the statistic V is equal to the right hand side of (45').

Denote by S the $2n + 1$-dimensional surface in the $2n + 4$-dimensional space of the variables $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ defined by the following equations

$$\sum_{\beta=1}^{n+2} u_\beta^2 = \sum_{\beta=1}^{n+2} v_\beta^2 = 1; \qquad \sum_{\beta=1}^{n+2} u_\beta v_\beta = 0. \tag{52}$$

denote by C the $2n + 1$-dimensional volume of the surface S, i.e.

$$C = \int_S dS. \tag{53}$$

Now we will assume that $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ are random variables and the joint probability distribution function is defined as follows: the point $(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})$ is restricted to points of S and the probability density function of S is defined by

$$\frac{dS}{C}. \tag{54}$$

Hence for any subset A of S the probability of A is equal to the $2n + 1$-dimensional volume of A divided by the $2n + 1$-dimensional volume of S. It should be remarked that the probability density function (54) is identical with the probability density function we would obtain if we were to assume that $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ are independently, normally distributed with zero means and unit variances and calculate the conditional density function under the restriction that $(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})$ is a point of S.

LEMMA 5. *The probability distribution of $\bar{V}$ defined in* (41), *calculated under the assumption that the joint probability density of the variables $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ $t_{i\beta}$ ($i = 1, \cdots, p$; $\beta = 1, \cdots, n + 2$) is given by the product of* (54) *and* (42), *is the same as the distribution of the statistic V calculated under the assumption that the variables $t_{i\beta}$ have the joint probability density function given in* (34).

Lemma 5 is an immediate consequence of lemma 3.

LEMMA 6. *Let L_p be an arbitrary p-dimensional flat space in the $n + 2$ dimensional Cartesian space, and let M_p be the flat space determined by the first p coordinate axes. Assuming that the joint probability density function of u_β, v_β, $t_{i\beta}$ ($i = 1, \cdots, p$; $\beta = 1, \cdots, n + 2$) is given by the product of* (54) *and* (42), *the conditional distribution of $\bar{V}$ calculated under the restriction that the points $R_1, \cdots, R_p$ lie in L_p, is the same as the conditional distribution of $\bar{V}$ calculated under the restriction that the points $R_1, \cdots, R_p$ lie in M_p. The point R_i denotes the point with the coordinates $t_{i1}, \cdots, t_{i,n+2}$.*

PROOF. Let P be the point with the coordinates $u_1, \cdots, u_{n+2}$ and let Q be the point with the coordinates $v_1, \cdots, v_{n+2}$. Let us rotate the coordinate system so that the first p axes lie in the flat space L_p. Denote the coordinates of P in the new system by $u'_1, \cdots, u'_{n+2}$, those of Q by $v'_1, \cdots, v'_{n+2}$, and those of R_i by $t'_{i1}, \cdots, t'_{i,n+2}$ ($i = 1, \cdots, p$). Let S' be the surface defined by

$$\Sigma u_\beta'^2 = \Sigma v_\beta'^2 = 1 \quad \text{and} \quad \Sigma u'_\beta v'_\beta = 0. \tag{55}$$

It is clear that the surface S' is identical with the surface S defined in (52). It is furthermore clear that if the joint density function of $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ is given by $\frac{dS}{C}$, the joint density function of $u'_1, \cdots, u'_{n+2}, v'_1, \cdots, v'_{n+2}$ is the same, i.e. it is given by $\frac{dS'}{C}$. It can readily be seen that for any given set of values $u'_1, \cdots, u'_{n+2}, v'_1, \cdots, v'_{n+2}$ the conditional joint probability density of the variates $t'_{i\beta}$ is given by the function obtained from (42) by substituting

$t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β, provided that for any given set of values $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ the joint conditional distribution of the variates $t_{i\beta}$ is given by (42). Hence, if the joint distribution of $u_1, \cdots, u_{n+2}$, $v_1, \cdots, v_{n+2}$ and $t_{i\beta}$ ($i = 1, \cdots, p; \beta = 1, \cdots, n+2$) is given by the product of (54) and (42), the joint probability density function of the variates u'_β, v'_β, $t'_{i\beta}$ ($i = 1, \cdots, p; \beta = 1, \cdots, n+2$) is obtained from that of u_β, v_β, $t_{i\beta}$ by substituting S' for S and $t'_{i\beta}$ for $t_{i\beta}$.

According to Lemma 4′, $\bar{V}$ can be expressed as a function of the angles θ_1, θ_2 and θ_3 defined in Lemma 4′. Each angle θ_k ($k = 1, 2, 3$) can be expressed as a function of the variables $t_{i\beta}$, u_β, v_β. It is obvious that the value of θ_k remains unchanged if we substitute $t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β. Hence also the value of $\bar{V}$ remains unchanged if we substitute $t'_{i\beta}$ for $t_{i\beta}$, u'_β for u_β and v'_β for v_β. Lemma 6 is a consequence of this fact and of the fact that the joint probability density of the variates $t'_{i\beta}$, u'_β and v'_β is identical with that of the variates $t_{i\beta}$, u_β and v_β.

LEMMA 7. *Assuming that the joint probability distribution of the variates* u_β, v_β, $t_{i\beta}$ ($i = 1, \cdots, p; \beta = 1, \cdots, n+2$) *is given by the product of* (54) *and* (42), *the conditional joint probability distribution of* $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$, *calculated under the restriction that the points* $R_i = (t_{i1}, \cdots, t_{i,n+2})$ ($i = 1, \cdots, p$) *lie in the flat space determined by the first* p *coordinate axes, is given by*

$$\frac{e^{-\frac{1}{2}\sum_{\gamma=p+1}^{n+2}\sum_{i=1}^{p}(\rho_i u_\gamma+\zeta_i v_\gamma)^2} f(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})\, dS}{\int_S e^{-\frac{1}{2}\sum_{\gamma=p+1}^{n+2}\sum_{i=1}^{p}(\rho_i u_\gamma+\zeta_i v_\gamma)^2} f(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})\, dS}, \tag{56}$$

where S *denotes the surface defined in* (52), *and* $f(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})$ *denotes the expected value of*

$$\left\|\begin{matrix} r_{11} & \cdots & r_{1p} \\ r_{21} & \cdots & r_{2p} \\ \vdots & & \vdots \\ r_{p1} & \cdots & r_{pp} \end{matrix}\right\|^{\frac{n+2-p}{2}} \qquad \left(r_{ij} = \sum_{\alpha=1}^{p} t_{i\alpha} t_{j\alpha}\right) \tag{57}$$

calculated under the assumption that the joint distribution of the variates $t_{i\beta}$ *is given by* (42).

PROOF. Denote by $\bar{R}_i$ the projection of R_i on the flat space determined by the first p coordinate axes, i.e. $\bar{R}_i = (t_{i1}, \cdots, t_{ip}, 0, \cdots, 0)$. Let l_1 be the length of $\bar{R}_1$, and let l_i be the distance of $\bar{R}_i$ from the flat space determined by the vectors $0\bar{R}_1, \cdots, 0\bar{R}_{i-1}$ ($i = 2, \cdots, p$). Then, as is known,

$$l_1 l_2 \cdots l_i = \sqrt{\left|\begin{matrix} r_{11} & \cdots & r_{1i} \\ r_{21} & \cdots & r_{2i} \\ \cdot & \cdots & \cdot \\ r_{i1} & \cdots & r_{ii} \end{matrix}\right|} \qquad (i = 1, \cdots, p), \tag{58}$$

where $r_{kl} = \sum_{\alpha=1}^{p} t_{k\alpha} t_{l\alpha}$.

We introduce the new variables

$$t^*_{i\gamma} = \frac{t_{i\gamma}}{l_i} \qquad (i = 1, \cdots, p; \gamma = p + 1, \cdots, n + 2). \tag{59}$$

Then the joint probability density function of the variates u_β, v_β, $t_{i\alpha}$, $t^*_{i\gamma}$ $(i = 1, \cdots, p; \beta = 1, \cdots, n + 2, \alpha = 1, \cdots, p, \gamma = p + 1, \cdots, n + 2)$ is given by

$$\frac{(l_1 \cdots l_p)^{n+2-p}}{C(2\pi)^{p(n+2)/2}} e^{-\frac{1}{2}\left[\sum_{i=1}^{p} \sum_{\alpha=1}^{p} (t_{i\alpha}-\rho_i u_\alpha - \zeta_i v_\alpha)^2 + \sum_{i=1}^{p} \sum_{\gamma=p+1}^{n+2} (l_i t^*_{i\gamma} - \rho_i u_\gamma - \zeta_i v_\gamma)^2\right]} \times \left(\prod_i \prod_\alpha dt_{i\alpha}\right)\left(\prod_i \prod_\gamma dt^*_{i\gamma}\right) dS. \tag{60}$$

Substituting zero for $t^*_{i\gamma}$ $(i = 1, \cdots, p, \gamma = p + 1, \cdots, n + 2)$ in (60), we obtain an expression which is proportional to the conditional joint probability density of the variates u_β, v_β, $t_{i\alpha}$ $(\beta = 1, \cdots, n + 2; i = 1, \cdots, p, \alpha = 1, \cdots, p)$, calculated under the restriction that the points R_i $(i = 1, \cdots, p)$ fall in the flat space determined by the first p coordinate axes. Hence this conditional density function is given by

$$A e^{-\frac{1}{2}\sum_{\gamma=p+1}^{n+2} \sum_{i=1}^{p} (\rho_i u_\gamma + \zeta_i v_\gamma)^2} (l_1 l_2 \cdots l_p)^{n+2-p} \times e^{-\frac{1}{2}\left[\sum_{i=1}^{p} \sum_{\alpha=1}^{p} (t_{i\alpha} - \rho_i u_\alpha - \zeta_i v_\alpha)^2\right]} dS \prod_i \prod_\alpha dt_{i\alpha} \tag{61}$$

where A denotes a constant. The conditional distribution of the variates u_β, v_β $(\beta = 1, \cdots, n + 2)$ is obtained from (61) by integrating it with respect to the variables $t_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, p)$. Because of (58), we see that the resulting formula is identical with (56). Hence Lemma 7 is proved.

LEMMA 8. *Let* $m_1 = u_1^2 + \cdots + u_p^2$; $m_2 = v_1^2 + \cdots + v_p^2$, *and* $m_3 = u_1 v_1 + \cdots + u_p v_p$. *If the joint distribution of the variates* $u_1, \cdots, u_{n+2}$, $v_1, \cdots, v_{n+2}$ *is given by* (54), *then the joint distribution of* m_1, m_2, m_3 *is given by*

$$\frac{B}{\sqrt{m_1 m_2 (1 - m_1)(1 - m_2)}} F_p(m_1) F_p(m_2) \Phi_p\left(\frac{m_3}{\sqrt{m_1 m_2}}\right) F_{n+2-p}(1 - m_1) \times F_{n+2-p}(1 - m_2) \Phi_{n+2-p}\left(\frac{-m_3}{\sqrt{(1 - m_1)(1 - m_2)}}\right) dm_1\, dm_2\, dm_3 \tag{62}$$

where B *denotes a constant,*

$$F_k(t) = \frac{1}{2^{k/2}\Gamma\left(\frac{k}{2}\right)} (t)^{(k-2)/2} e^{-\frac{1}{2}t} \quad \text{and} \quad \Phi_k(t) = \frac{\Gamma\left(\frac{k}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{k-1}{2}\right)} (1 - t^2)^{(k-3)/2}. \tag{63}$$

PROOF. Let $m_1' = u_{p+1}^2 + \cdots + u_{n+2}^2$, $m_2' = v_{p+1}^2 + \cdots + v_{n+2}^2$,

$m_3' = u_{p+1}v_{p+1} + \cdots + u_{n+2}v_{n+2}$, $\bar{m}_3 = \dfrac{m_3}{\sqrt{m_1 m_2}}$ and $\bar{m}_3' = \dfrac{m_3'}{\sqrt{m_1' m_2'}}$. First we calculate the joint distribution of m_1, m_2, $\bar{m}_3$ m_1', m_2', $\bar{m}_3'$ under the assumption that $u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2}$ are normally independently distributed with zero means and unit variances. This joint distribution is given by

$$(64) \quad \begin{aligned} &F_p(m_1)F_p(m_2)\Phi_p(\bar{m}_3)F_{n+2-p}(m_1')F_{n+2-p}(m_2') \\ &\qquad \times \Phi_{n+2-p}(\bar{m}_3')\, dm_1\, dm_2\, d\bar{m}_3\, dm_1'\, dm_2'\, d\bar{m}_3'. \end{aligned}$$

Hence the joint distribution of m_1, m_2, m_3, m_1', m_2', m_3' is given by

$$(65) \quad \begin{aligned} &\frac{1}{\sqrt{m_1 m_2 m_1' m_2'}} F_p(m_1)F_p(m_2)\Phi_p\left(\frac{m_3}{\sqrt{m_1 m_2}}\right) F_{n+2-p}(m_1')F_{n+2-p}(m_2') \\ &\qquad \times \Phi_{n+2-p}\left(\frac{m_3'}{\sqrt{m_1' m_2'}}\right) dm_1\, dm_2\, dm_3\, dm_1'\, dm_2'\, dm_3'. \end{aligned}$$

The required conditional distribution of m_1, m_2, m_3 is equal to the conditional distribution of m_1, m_2, m_3 obtained from the joint distribution (65) under the restrictions $m_1 + m_1' = 1$, $m_2 + m_2' = 1$ and $m_3 + m_3' = 0$. Hence if in (65) we substitute $1 - m_1$ for m_1', $1 - m_2$ for m_2' and $-m_3$ for m_3' we obtain an expression proportional to the conditional distribution of m_1, m_2, m_3. This proves Lemma 8.

LEMMA 9. *For any point* $(u_1, \cdots, u_{n+2}, v_1, \cdots, v_{n+2})$ *of the surface S defined in* (52) *the expected value of* (57) (*calculated under the assumption that* (42) *is the joint distribution of* $t_{i\beta}$) *is a function of* m_1 m_2, *and* m_3 *only, where* m_1, m_2 *and* m_3 *are defined in Lemma 8.*

PROOF. Let $\| \lambda_{\alpha\beta} \|$ $(\alpha, \beta = 1, \cdots, p)$ be an orthogonal matrix such that

$$(66) \qquad \lambda_{1\beta} = \frac{u_\beta}{\sqrt{u_1^2 + \cdots + u_p^2}} \qquad (\beta = 1, \cdots, p)$$

and

$$(67) \qquad \lambda_{2\beta} = \frac{u_\beta + \lambda v_\beta}{\sqrt{\sum_{\beta=1}^{p} (u_\beta + \lambda v_\beta)^2}} \qquad (\beta = 1, \cdots, p)$$

where

$$\lambda = \frac{-\sum_1^p u_\beta^2}{\sum_1^p u_\beta v_\beta}.$$

Let

$$(68) \qquad t_{i\alpha}' = \sum_{\beta=1}^{p} \lambda_{\alpha\beta} t_{i\beta} \qquad (\alpha = 1, \cdots, p).$$

Then the variates $t'_{i\alpha}$ are independently and normally distributed with unit variances. Since for any point of S, $E(t_{i\alpha}) = \rho_i u_\alpha + \zeta_i v_\alpha$, we have because of (66), (67) and (68)

$$E(t_{i\gamma}) = 0 \qquad (i = 1, \cdots, p, \gamma = 3, 4, \cdots, p),$$
$$E(t_{i1}) = \varphi_{i1}(m_1, m_2, m_3),$$
and
$$E(t_{i2}) = \varphi_{i2}(m_1, m_2, m_3).$$

Hence the joint distribution of the variates $t'_{i\alpha}$ $(i = 1, \cdots, p; \alpha = 1, \cdots, p)$ depends merely on m_1, m_2 and m_3. Since $r_{ij} = \sum_{\alpha=1}^{p} t_{i\alpha}t_{j\alpha} = \sum_{\alpha=1}^{p} t'_{i\alpha}t'_{j\alpha}$, the expression (57) can be expressed as a function of the variables $t'_{i\alpha}$. Hence the distribution of the expression (57) depends merely on the parameters m_1, m_2, and m_3. This proves Lemma 9.

The main result of this section is the following

THEOREM. *Let V be the statistic given in* (25) *and let the joint distribution of the variates $t_{i\beta}$ $(i = 1, \cdots, p; \beta = 1, \cdots, n + 2)$ be given by* (34). *Then the probability distribution of V is the same as the distribution of*

$$(69) \qquad -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)}$$

where the joint distribution of m_1, m_2 and m_3 is equal to a constant multiple of the product of the following three factors: the expression (62), *the exponential $e^{\frac{1}{2}(m_1\sum\rho_i^2 + 2m_3\sum\rho_i\zeta_i + m_2\sum\zeta_i^2)}$ and the expected value of*

$$(70) \qquad \left(\begin{vmatrix} r_{11} & \cdots & r_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ r_{p1} & \cdots & r_{pp} \end{vmatrix}\right)^{(n+2-p)/2} \qquad \left(r_{ij} = \sum_{\alpha=1}^{p} t_{i\alpha} t_{j\alpha}\right).$$

The expected value of (70) *is calculated under the assumption that the variates $t_{i\alpha}$ are normally and independently distributed with unit variances and $E(t_{i\alpha}) = \rho_i u_\alpha + \zeta_i v_\alpha$ $(i = 1, \cdots, p; \alpha = 1, \cdots, p)$ where $\sum_{\alpha=1}^{p} u_\alpha^2 = m_1$, $\sum_{\alpha=1}^{p} v_\alpha^2 = m_2$ and $\sum_{\alpha=1}^{p} v_\alpha u_\alpha = m_3$. The domain of the variables m_1, m_2 and m_3 is given by the inequalities: $0 \leq m_1 \leq 1; 0 \leq m_2 \leq 1; -\sqrt{m_1m_2} \leq m_3 \leq \sqrt{m_1m_2}$.*

PROOF. First we note that the expected value of (70) is a function of m_1, m_2 and m_3 only. Let P be the point with the coordinates $u_1, \cdots, u_{n+2}$, and Q the point with the coordinates $v_1, \cdots, v_{n+2}$. Assume that the points $R_i = (t_{i1}, \cdots, t_{i,n+2})$ $(i = 1, \cdots, p)$ lie in the flat space determined by the first p coordinate axes. Assume furthermore that $u_1v_1 + \cdots + u_{n+2}v_{n+2} = 0$ and that the lengths of the vectors $0P$ and $0Q$ are equal to 1. Then

$$\cos\theta_1 = \sqrt{u_1^2 + \cdots + u_p^2}; \qquad \cos\theta_2 = \sqrt{v_1^2 + \cdots + v_p^2}$$

and

$$\cos\theta_3 = \frac{u_1v_1 + \cdots + u_pv_p}{\sqrt{u_1^2 + \cdots + u_p^2}\sqrt{v_1^2 + \cdots + v_p^2}},$$

where θ_1 denotes the angle between $0P$ and the flat space L_p determined by the vectors $0R_1, \cdots, 0R_p$; θ_2 denotes the angle between $0Q$ and L_p, and θ_3 denotes the angle between the projections of $0P$ and $0Q$ on L_p. According to Lemma 4′ the statistic $\bar{V}$ defined in (41) is equal to

$$\bar{V} = -n \frac{\cos\theta_1 \cos\theta_2 \cos\theta_3}{\cos^2\theta_1 \cos^2\theta_2 \cos^2\theta_3 - \sin^2\theta_1 \sin^2\theta_2} = -n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)} \tag{71}$$

where

$$m_1 = \cos^2\theta_1 = u_1^2 + \cdots + u_p^2, \qquad m_2 = \cos^2\theta_2 = v_1^2 + \cdots + v_p^2 \tag{72}$$
$$\text{and} \quad m_3 = \cos\theta_1 \cos\theta_2 \cos\theta_3 = u_1v_1 + \cdots + u_pv_p .$$

It follows from Lemmas 5 and 6 that the distribution of V is the same as the conditional distribution of $\bar{V}$ calculated under the assumption that the unconditional joint probability density of the variates u_β, v_β and $t_{i\beta}$ is given by the product of (54) and (42) and under the restriction that the points R_i $(i = 1, \cdots, p)$ fall in the flat space determined by the first p coordinate axes. Since $e^{-\frac{1}{2}\sum_{\gamma=p+1}^{n+2}\sum_{i=1}^{p}(\rho_i u_\gamma + \zeta_i v_\gamma)^2}$ is a constant multiple of

$$e^{\frac{1}{2}(m_1\Sigma\rho_i^2 + 2m_3\Sigma\rho_i\zeta_i + m_2\Sigma\zeta_i^2)} \tag{73}$$

from Lemmas 7, 8 and 9 it follows readily that the joint conditional distribution of $m_1 = u_1^2 + \cdots + u_p^2$, $m_2 = v_1^2 + \cdots + v_p^2$ and $m_3 = u_1v_1 + \cdots + u_pv_p$ is equal to a constant multiple of the product of (62), (73) and the expected value of 70. This proves our theorem.

It can be shown that the variates m_1, m_2 and m_3 are of the order $\frac{1}{n}$ in the probability sense. Hence

$$-n \frac{m_3}{m_3^2 - (1 - m_1)(1 - m_2)} = nm_3(1 + \epsilon) \tag{74}$$

where ϵ is of the order $\frac{1}{n}$. Hence we can say: *even for moderately large n the distribution of the statistic V is well approximated by the distribution of* nm_3, *where the joint distribution of* m_1, m_2 *and* m_3 *is equal to a constant multiple of the product of* (62), (73) *and the expected value of* (70).

If $n + 2 - p$ is an even integer, the expected value of (70) is obviously an elementary function of m_1, m_2 and m_3. Hence, if $n + 2 - p$ is even, the joint distribution of m_1, m_2 and m_3 is also an elementary function of m_1, m_2 and m_3.

If the constants ρ_i and ζ_i $(i = 1, \cdots, p)$ in formula (34) are equal to zero, the expected value of (70) is a constant and the joint distribution of m_1, m_2 and m_3 is given by (62).

ASYMPTOTIC DISTRIBUTION OF RUNS UP AND DOWN[1]

By J. Wolfowitz

Columbia University

1. Introduction. Let $a_1, a_2, \cdots, a_n$ be any n unequal numbers and let $S = (h_1, h_2, \cdots, h_n)$ be a random permutation of them, with each permutation having the same probability, which is therefore $\frac{1}{n!}$. Let R be the sequence of signs ($+$ or $-$) of the differences $h_{i+1} - h_i$ $(i = 1, 2, \cdots, n-1)$. Then R is also a chance variable. A sequence of p successive $+$ $(-)$ signs not immediately preceded or followed by a $+$ $(-)$ sign is called a run up (down) of length p. The term "run" applies to both runs up and runs down. As an example, if $S = (4\ 6\ 2\ 3\ 5)$, then in $R = (+ - + +)$ there are three runs, one up of length one, one down of length one, and one up of length two.

The purpose of this paper is to establish several theorems about the limiting distributions of a class of functions of runs up and down. These results are applicable to certain techniques which have been employed in quality control and the analysis of economic time series. They are also shown to apply to a large class of "runs."

2. Joint distribution of runs of several lengths. Let r_p be the number of runs of length p in R and r'_p the number of runs of length p or more in R. Then r_p and r'_p are chance variables. The expectations $E(r_p)$ and $E(r'_p)$, the variances $\sigma^2(r_p)$ and $\sigma^2(r'_p)$, and the covariances $\sigma(r_{p_1}r_{p_2})$ are given by Levene and Wolfowitz [1]. They are all of the order n. Let

$$y_p = \frac{r_p - E(r_p)}{\sqrt{n}},$$

$$y'_p = \frac{r'_p - E(r'_p)}{\sqrt{n}}.$$

Our first results are embodied in the following theorem:

Theorem 1. *Let l be any non-negative integer. The joint distribution of $y_1, \cdots, y_l, y'_{(l+1)}$, approaches the normal distribution as $n \to \infty$.*

We shall give the proof for the case $l = 1$, but it will easily be seen to be perfectly general.

Let $x_{pi} = 1$ if the sign ($+$ or $-$) of $h_{i+1} - h_i$ is the initial sign of a run of length p, and let $x_{pi} = 0$ otherwise. Let $w_{pi} = 1$ if the sign of $h_{i+1} - h_i$ is the

---

[1] Part of the results of this paper was presented to the Institute of Mathematical Statistics and the American Mathematical Society at their joint meeting in New Brunswick, N. J., on September 13, 1943.

initial sign of a run of length p or more, and let $w_{pi} = 0$ otherwise. Let $x_{pn} = w_{pn} = 0$. Then

$$r_1 = \sum_{i=1}^{n} x_{1i},$$

$$r_2' = \sum_{i=1}^{n} w_{2i}.$$

Now write $\alpha = n^{3/4}$, $\beta = n^{1/4}$, and consider the β sequences

$$h_{(j-1)\alpha+1},\ h_{(j-1)\alpha+2},\ \cdots,\ h_{j\alpha} \qquad (j = 1, 2, \cdots, \beta).$$

(Strictly speaking, we should employ the largest integer in α. Since what is meant is clear and since we are dealing with an asymptotic property, we shall omit this useless nicety.) Let x'_{pi} and w'_{pi} have the same definitions relative to each of these sequences that x_{pi} and w_{pi} have relative to the sequence S. The accented and unaccented x's and w's are not always the same, because the partitioning of the sequence S sometimes breaks up runs and creates others. Thus we might have $x_{p\alpha} = 1$, but $x'_{p\alpha}$ always $= 0$.

It is easy to see that there exists a positive number d such that

$$\sum_{i=1}^{n} | x_{1i} - x'_{1i} | < d\beta,$$

$$\sum_{i=1}^{n} | w_{2i} - w'_{2i} | < d\beta.$$

If, therefore, we define

$$z_1 = \frac{\sum_{i=1}^{n} [x'_{1i} - E(x'_{1i})]}{\sqrt{n}},$$

$$z_2' = \frac{\sum_{i=1}^{n} [w'_{2i} - E(w'_{2i})]}{\sqrt{n}},$$

we have

$$| z_1 - y_1 | < \frac{2\, d\beta}{\sqrt{n}}$$

$$| z_2' - y_2' | < \frac{2\, d\beta}{\sqrt{n}}$$

and

$$\frac{d\beta}{\sqrt{n}} \to 0.$$

Hence, if the joint limiting distribution of z_1 and z_2' is normal, so is that of y_1 and y_2'.

The chance variables

$$r_{1j} = \sum_{i=(j-1)\alpha+1}^{j\alpha} x_{1i}'$$

$$r_{2j}' = \sum_{i=(j-1)\alpha+1}^{j\alpha} w_{2i}' \qquad (j = 1, 2, \cdots, \beta)$$

have the same joint distribution for all values of j. For x_{1i}' and w_{2i}', $((j-1)\alpha < i < j\alpha)$, depend only on the relative magnitude of the elements of the sequence

$$h_{(j-1)\alpha+1}, \cdots, h_{j\alpha},$$

not upon the particular values which the elements take, and all permutations of the sequence have equal probability. Clearly r_{1j} and r_{2j}' are independent, in the probability sense, of $r_{1j'}$ and $r_{2j'}'$ $(j \neq j')$, because of the definitions of x_{1i}' and w_{2i}'. (However, r_{1j} and r_{2j}' are not independent, because x_{1i}' and w_{2i}' cannot both be 1.) From the results of [1] it follows that for sufficiently large n the absolute value of the correlation coefficient between r_{1j} and r_{2j}' is less than a number smaller than 1. By the methods of [1] it can easily be shown that the ratio of the fourth order moments of r_{1j} and r_{2j}' about their means to the square of the variance of either, is bounded for sufficiently large n. Hence by Liapounoff's theorem (see, for example, Cramer [2], Uspensky [5]), z_1 and z_2' are jointly normally distributed in the limit. Hence so are y_1 and y_2' and the theorem is proved.

3. Generalization of Theorem 1. Examination of the proof of Theorem 1 shows that it rests on the following two properties of runs up and down:

a) Partition of the sequence S into subsequences affects at most d runs in each sub-sequence, where d is a fixed positive number independent of n.

b) After partition the totals of runs of each length in any sub-sequence (the definition now relates to the subsequence) are independent in the probability sense of the totals of runs in any other subsequence, and satisfy some condition (such as the Liapounoff) sufficient to make the components of the sum of the vectors jointly normally distributed in the limit.

Hence if we adopt other definitions of runs which meet conditions (a) and (b) above, the total numbers of each of these various kinds of runs will be in general jointly asymptotically normally distributed. For example, if s_p and s_p' be the numbers of runs *up* of length p and of length p or more, respectively, and if t_p and t_p' are the same quantities referring to runs *down*, then, with l and k any positive integers,

$$s_1, s_2, \cdots, s_l, \qquad s_{(l+1)}', \qquad t_1, t_2, \cdots, t_k$$

are jointly asymptotically normally distributed. However, if $t'_{(k+1)}$ is included in this set, since

$$s'_1 = s_1 + s_2 + \cdots + s_l + s'_{(l+1)}$$

and

$$t'_1 = t_1 + t_2 + \cdots + t_k + t'_{(k+1)}$$

differ by at most one, the limiting distribution is degenerate, i.e., its covariance matrix is only semi-definite.

As another example, if we define a bizarre run as, say, the occurrence of a run up of length 5, followed, 17 elements later, by a run down of length 14, then the number of runs of this type is asymptotically normally distributed with expectation and variance of order n.

4. Additive functions of runs of all lengths. Combining the numbers of runs of all lengths greater than a given length generally involves a loss of information. The following theorem on additive functions of runs up and down may be of general interest and of utility in avoiding this undesirable situation.

THEOREM 2. *Let $f(i)$ be a function, defined for all positive integral values of i, which fulfills the following conditions:*

a) *There exists a pair of positive integers, a and b, such that*

$$\frac{f(a)}{f(b)} \neq \frac{a}{b} \tag{4.1}$$

b) *for any $\epsilon_1 > 0$ there exists a positive integer $N(\epsilon_1)$ such that, for all $n \geq N(\epsilon_1)$,*

$$\sum_{i=N(\epsilon_1)}^{i=n-1} | f(i) | \sigma(r_i) < \epsilon_1 n \tag{4.2}$$

where n, of course, has the same meaning as in the preceding sections. Let $F(S)$, a function of the chance sequence S, be defined as follows:

$$F(S) = \sum_{i=1}^{(n-1)} f(i) r_i . \tag{4.3}$$

Then the distribution of $\dfrac{F(S) - E[F(S)]}{\sigma[F(S)]}$ *approaches the normal distribution as* $n \to \infty$.

As an example, let $f(i) \equiv 1$. Then $F(S) \equiv r'_1$, whose limiting distribution is normal by Theorem 1.

This theorem is the exact analogue of Theorem 2 of [3] and the proof of the latter carries over without difficult changes except in one important respect. A difficulty in the proof of the theorem in [3] lay in proving Lemma 4, and this lemma has to be proved completely anew. We shall limit ourselves here to doing just that. Lemma 2 of Theorem 2 of [3], whose only role was to help in proving Lemma 4, has no analogue in our present problem, but all the others do. It will therefore be sufficient if we prove the following:

LEMMA. *There exists a constant $c > 0$, such that, for all n sufficiently large,*

(4.4) $$\sigma^2[F(S)] > cn.$$

Condition (a) of the theorem is imposed simply in order that the result be not trivial. For, if (a) does not hold, we have that

$$f(i) \equiv if(1),$$

and

$$F(S) \equiv f(1) \Sigma ir_i$$
$$\equiv (n-1)f(1) = \text{a constant}.$$

Suppose that

$$f(i) \equiv ui + v,$$

with u and v constants, and $v \neq 0$. Then by Theorem 1

$$F(S) = u(n-1) + vr_1'$$
$$= vr_1' + \text{a constant}$$

is asymptotically normally distributed with variance of order n. Without loss of generality we may therefore assume that

(4.5) $$f(i) \not\equiv ui + v.$$

From this it follows that there exists an integer $A \geq 2$ such that

(4.6) $$f(A-1) + f(A+1) \neq 2f(A).$$

Our object is to prove that $\sigma^2[F(S)]$ is at least of order n. The basic idea of the proof will be to construct two sets, say L_1 and L_2, of sequences S, such that the (same) probability of each is not less than a positive lower bound independent of n, and such that there exists a one-to-one correspondence between the sequences of L_1 and those of L_2 so that, if S_1 is a member of L_1 and S_2 the corresponding sequence in L_2,

$$|F(S_1) - F(S_2)| \geq g\sqrt{n},$$

where g is a positive constant independent of n. It is easy to see that such a construction would prove the lemma.

We shall call the subsequence $(h_i, h_{i+1}, \cdots, h_{i+2A})$ of S, a run of type T_1 or simply a run T_1 (the notion will be used only for the proof of this lemma) if the following conditions are fulfilled:

(4.7) each of the signs of $(h_{i+1} - h_i)$ and $(h_{i+A+1} - h_{i+A})$ is the initial sign of a run of length A.

(4.8) if $i \neq 1$, the sign of $(h_i - h_{i-1})$ is not the final sign of a run of length A.

(4.9) if $i + 2A \neq n$, the sign of $(h_{i+2A+1} - h_{i+2A})$ is not the initial sign of a run of length A.

(4.10) after the transformation H, which interchanges h_{i+A-1} and h_{i+A}, has operated on the run, the sign of $(h_{i+1} - h_i)$ is the initial sign of a run of length $A - 1$, and the sign of $(h_{i+A-1} - h_{i+A})$, in the new ordering, is the initial sign of a run of length $A + 1$.

Thus, with $A = 2$ and $n = 7$, if $S = (7145326)$, then $R = (- + + - - +)$, and (1 4 5 3 2) is a run T_1, for after the transformation H has been applied we have (1 5 4 3 2) which gives $(+ - - -)$. The result of the operation H on a run T_1 will be called a run T_2.

The number r^* of runs T_1 and the number r^{**} of runs T_2 each have expected values and variances of order n, by considerations similar to those of [1]. Hence, for an arbitrarily small positive ϵ there exists a positive constant q such that, for all n sufficiently large, the probability $P\{r^* + r^{**} \geq qn\}$ of the set L^* of sequences S which satisfy the relation in braces, is not less than $1 - \epsilon$.

The set L^* can be divided into disjunct sets (families) as follows: Let $S(0)$ be any sequence S in L^* which has no runs T_2 (any doubt about the existence of such sequences will be soon removed) and let $r^*[S(0)] = m$. Hence $m \geq qn$. Operating with the transformation H on each of the m runs T_1 of $S(0)$ we get a set $S(1)$ of m different sequences for each of which $r^* = m - 1, r^{**} = 1$. Operating again with H on each of the pairs of runs T_1 of the sequence $S(0)$ we get a set $S(2)$ of $\binom{m}{2}$ distinct sequences for each of which $r^* = m - 2, r^{**} = 2$, etc. The process stops with $S(m)$, which contains a single sequence, for which $r^* = 0$, $r^{**} = m$. The set $S(i)$ contains $\binom{m}{i}$ different sequences for each of which $r^* = m - i, r^{**} = i$. The union of the sets $S(i)$ $(i = 1, 2, \cdots, m)$ will be called the family whose generator is $S(0)$. The sets $S(i)$ are obviously disjunct. Any sequence S in L^* belongs to one and only one family. For if we operate on *all* of its runs T_2 with H (which is its own inverse), we obtain the generator of the family to which it belongs. This also proves the existence of sequences in L^* for which $r^{**} = 0$.

Consider any family F whose generator is a sequence for which $r^* = m \geq qn$. It is easy to see that, when n is sufficiently large, the ratio of the total number of sequences S in the sets L_1^* and L_2^*, where

$$L_1^* = \sum_{i=0}^{i=\frac{1}{2}(m-\sqrt{m})} S(i),$$

and

$$L_2^* = \sum_{i=\frac{1}{2}(m+\sqrt{m})}^{i=m} S(i),$$

to the total number of sequences in F is greater than a fixed positive constant K'.

We are now ready to construct L_1 and L_2. The set L_1 is the union of the sets L_1^* of all the families in L^*, and the set L_2 is the union of the sets L_2^* of all the families in L^*. The probability of L_1 and of L_2 is therefore not less than $\frac{1}{2}K'(1 - \epsilon)$. The one-to-one correspondence is effected as follows: The subset $S\left(\frac{m}{2} - \frac{\sqrt{m}}{2} - j\right)$ of the set L_1^* of any family is to correspond to the

subset $S\left(\frac{m}{2} + \frac{\sqrt{m}}{2} + j\right)\left(j = 0, 1, 2, \cdots, \frac{m}{2} - \frac{\sqrt{m}}{2}\right)$ of the set L_2^* of the same family. The individual sequences of either of the two subsets may be made to correspond to those of the other in any manner whatsoever. Any sequence S_1 in L_1 and its corresponding sequence S_2 in L_2 thus differ only in the numbers of runs T_1 and T_2, but are identical in the numbers of all other runs. They differ in at least $\sqrt{m}$ runs. Hence,

$$\begin{aligned} | F(S_1) - F(S_2) | &\geq \sqrt{m} \, | 2f(A) - f(A - 1) - f(A + 1) | \\ &\geq \sqrt{qn} \, | 2f(A) - f(A - 1) - f(A + 1) | . \end{aligned}$$

This is the required result with

$$g = \sqrt{q} \, | 2f(A) - f(A - 1) - f(A + 1) | .$$

Hence the lemma and the theorem are proved.

The remarks of section 3 also apply to Theorem 2.

5. The distribution of long runs. Certain tests in use in quality control of manufactured products are based on the occurrence of long runs. Since the mean and variance of r_p, for any fixed p, are of order n, it follows that the probability that $r_p \neq 0$ approaches 1 (with increasing n). In order to base a test on the occurrence of a run of length p in long sequences it is therefore necessary to make p a function of n. This function must be a suitable one, because if p is, for example, of the order n, the probability that $r_p = 0$ approaches 1; p should, therefore, be neither too short nor too long.

The following theorem will help give the answer to this problem:

THEOREM 3. *Let p vary with n, so that*

$$\frac{(p + 1)!}{n} = \frac{1}{K}$$

with K a fixed positive number. Then

$$\lim_{n \to \infty} P\{r_p = j\} = e^{-2K} \frac{(2K)^j}{j!} \qquad (j = 0, 1, 2, \cdots)$$

i.e., r_p *has in the limit the Poisson distribution with mean* $2K$.

The proof will consist in showing that the moments of r_p approach the moments of a Poisson distribution with mean $2K$ as $n \to \infty$. This is sufficient (v. Mises [4]).

Let $x_i = 1$ if the sign of $h_{i+1} - h_i$ is the initial sign of a run of length p, and $x_i = 0$ otherwise. The probability that $x_i = 1$ is, by [1], Section [4], $\frac{2(p^2 + 3p + 1)}{(p + 3)!}$ for all i with a fixed number of exceptions.[2] Write $B = \frac{2}{(p + 1)!}$; then

$$P\{x_i = 1\} = B + o(B),$$

[2] Since these exceptions (at the ends of the sequence S) have no effect on the asymptotic theory, they will henceforth be ignored.

where the symbol $o(B)$ means that $\lim \frac{o(B)}{B} = 0$. Let y_i ($i = 1, 2, \cdots, n$) be independent chance variables with the same distribution: $P\{y_i = 1\} = B$, $P\{y_i = 0\} = 1 - B$. Then it is easy to see that $Y = \sum_{i=1}^{n} y_i$ has in the limit the Poisson distribution with mean $2K$ and that its moments approach the moments of the same Poisson distribution. Hence it will be sufficient to show that in the limit Y and r_p have the same moments.

If $q, \alpha_1, \alpha_2, \cdots, \alpha_q$ and $i_1 < i_2 < \cdots < i_q$ are positive integers, we have that

$$E(y_{i_1}^{\alpha_1} y_{i_2}^{\alpha_2} \cdots y_{i_q}^{\alpha_q}) = E(y_{i_1} y_{i_2} \cdots y_{i_q}) = \prod_{j=1}^{q} E(y_{i_j}) = B^q \tag{5.1}$$

and

$$0 \leq E(x_{i_1}^{\alpha_1} x_{i_2}^{\alpha_2} \cdots x_{i_q}^{\alpha_q}) = E(x_{i_1} x_{i_2} \cdots x_{i_q}). \tag{5.2}$$

Also

$$E(r_p^l) = E\left[\sum_{i=1}^{n} x_i\right]^l. \tag{5.3}$$

After expansion of the right member of (5.3), we may replace, in accord with (5.2), each of the non-zero exponents of the x's by 1. The same operation on the terms of the expansion of the right member of

$$E(Y^l) = E\left[\sum_{i=1}^{n} y_i\right]^l, \tag{5.4}$$

is valid in accord with (5.1).

Let $i_1 < i_2 < \cdots < i_q$. In the expression

$$E(x_{i_1} x_{i_2} \cdots x_{i_q}), \tag{5.5}$$

let q be the "weight." A subsequence of consecutive x's in (5.5) (it may consist of a single x) which is such that the indices of two consecutive x's differ by less than $(p + 3)$, while the subsequence cannot be expanded on either side without violating this requirement, will be called a "cycle." Let c be the number of cycles in (5.5). By [1], Section 4, if x_i and x_j are in different cycles, i.e., $|i - j| \geq (p + 3)$, then x_i and x_j are independently distributed. If, therefore, $q = c$, we have that

$$E(x_{i_1} x_{i_2} \cdots x_{i_q}) = \prod_{j=1}^{q} E(x_{i_j}) = B^q + o(B^q). \tag{5.6}$$

If $q > c = 1$, we have, also from [1], Section 4, that

$$E(x_{i_1} x_{i_2} \cdots x_{i_q}) \leq E(x_{i_1} x_{i_2}) = o(B). \tag{5.7}$$

If $q > c$ and if there are two indices in the expression (5.5) which differ by less than p, then

$$E(x_{i_1}x_{i_2} \cdots x_{i_q}) = 0. \tag{5.8}$$

For x_i and x_j cannot both initiate runs of length p if $| i - j | < p$.

Let us now return to the expansions of the right members of (5.3) and (5.4), in which the exponents have been replaced as described before. Let the weight and cycle definitions also apply to terms of the type

$$E(y_{i_1}y_{i_2} \cdots y_{i_q}). \tag{5.9}$$

From (5.1) and (5.6) it follows that, in the limit, the contributions to $E(r_p^l)$ and $E(Y^l)$ of the sums of those terms for which $q = c$, are the same. Let W and W' be the sums of all the remaining terms in $E(r_p^l)$ and $E(Y^l)$, respectively. If we can show that

$$\lim W = \lim W' = 0 \tag{5.10}$$

we will have proven that

$$\lim E(r_p^l) = \lim E(Y^l) \tag{5.11}$$

and with it the theorem.

Let $B = O[f(n)]$ mean, as usual, that $| B | \leq Mf(n)$ for all n and a fixed $M > 0$. The number of terms in W' with fixed q and c ($c < q$, by definition of W') is $O(n^c p^{q-c})$. From (5.1) the value of the sum of all such terms is $O(B^q n^c p^{q-c})$. Now

$$nB = O(1)$$

by the hypothesis of the theorem. From the definition of p,

$$p = o(n)$$

and hence

$$pB = o(1).$$

Therefore

$$\begin{aligned} B^q n^c p^{q-c} &= (nB)^c (pB)^{q-c} \\ &= o(1). \end{aligned}$$

Since $q \leq l$, there are only a fixed number of such sums. Hence $\lim W' = 0$.

The number of terms in W with fixed q and c ($c < q$) is $O(n^c p^{q-c})$. However, most of these are of the type in (5.8) and therefore vanish. Those which do not vanish are $O(n^c)$ in number. Since $q > c$ we have by application of (5.7) that each term is $o(B^c)$. Hence the value of the sum of these terms is $o(n^c B^c) = o(1)$. Since $q \leq l$, there are a fixed number of such sums. Hence $\lim W = 0$.

This proves (5.10) and with it the theorem.

It is possible to generalize this result in a manner similar to that of Section 3.

The author is obliged to W. Allen Wallis who first drew his attention to problems in runs up and down, and to Howard Levene, who read the manuscript of this paper.

REFERENCES

[1] H. Levene and J. Wolfowitz, *Annals of Math. Stat.*, Vol. 15 (1944).
[2] Harald Cramér, *Random Variables and Probability Distributions*, Cambridge, 1937.
[3] J. Wolfowitz, *Annals of Math. Stat.*, Vol. 13 (1942), p. 247.
[4] R. v. Mises, *Zeitschrift fuer die angewandte Math. und Mechanik*, Vol. 1 (1921), p. 298.
[5] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937.

STATISTICAL ANALYSIS OF CERTAIN TYPES OF RANDOM FUNCTIONS

By H. Hurwitz, Jr. and M. Kac

Cornell University

1. Introduction. In solving certain physical problems (Brownian movements, shot effect) one is often led to the study of superpositions of random pulses. More precisely, one is led to sums of the type

$$F(t) = \sum_{j=1}^{N} f(t - t_j), \tag{1}$$

where N and the t_j's are random variables and a function $P(t)$ is given such that $\int_\Delta P(t)\, dt$ represents the average number of pulses occurring during the time interval Δ.

We propose to give a fairly detailed treatment of those statistical properties of $F(t)$ which may be of interest to a physicist and at the same time pay careful attention to the mathematical assumptions which underly the applications. It may also be pointed out that our results could be applied to the theory of time series.

2. Statistical assumptions and the distribution of N. The statistical assumptions can be formulated as follows:

1. The t_j's form an infinite sequence of independent identically distributed random variables each having $p(t)$ as its probability density.
2. N is capable of assuming the values $0, 1, 2, 3, \cdots$ only, and N is independent of the t_j's.
3. If $M(\Delta; N)$ denotes the number of those t_j's among the first N, which fall within the interval Δ, then for non-overlaping intervals Δ_1 and Δ_2 the random variables $M(\Delta_1 ; N)$ and $M(\Delta_2 ; N)$ are independent.

We now state our first theorem.[1]

Theorem 1. *Assumptions* 1, 2, 3 *imply that N is distributed according to Poisson's law, i.e.*

$$\text{Prob}\,\{N = r\} = e^{-h}\frac{h^r}{r!},$$

where $h = \int_{-\infty}^{+\infty} P(t)\, dt$.

[1] For a different approach to Poisson's distribution see W. Feller, *Math. Ann.* 113 (1937) in particular pp. 113–160.

Our proof is based on considerations of characteristic functions. Let $\psi_\Delta(x)$ be 1 if x belongs to the interval Δ and 0 otherwise. Thus

$$M(\Delta; N) = \sum_{k=1}^{N} \psi_\Delta(t_k).$$

From the independence of $M(\Delta_1; N)$ and $M(\Delta_2; N)$ it follows that for every pair of real numbers ξ and η we have[2]

$$E\left[\exp\left\{i\left(\xi \sum_{k=1}^{N} \psi_{\Delta_1}(t_k) + \eta \sum_{k=1}^{N} \psi_{\Delta_2}(t_k)\right)\right\}\right] = E\left[\exp\left\{i\xi \sum_{k=1}^{N} \psi_{\Delta_1}(t_k)\right\}\right] E\left[\exp\left\{i\eta \sum_{k=1}^{N} \psi_{\Delta_2}(t_k)\right\}\right],$$

where $E[x]$ denotes the mathematical expectation, or mean value, of x. Letting $q(r) = \text{Prob}\,\{N = r\}$ and using first the independence of N and the t_j's and then the fact that the t_j's are independent and identically distributed we obtain

$$\sum_{r=0}^{\infty} q(r)(E[\exp\{i(\xi\psi_{\Delta_1}(t) + \eta\psi_{\Delta_2}(t))\}])^r = \sum_{r=0}^{\infty} q(r)(E[\exp\{i\xi\psi_{\Delta_1}(t)\}])^r \sum_{r=0}^{\infty} q(r)(E[\exp\{i\eta\psi_{\Delta_2}(t)\}])^r. \tag{2}$$

An easy calculation gives

$$E[\exp\{i\xi\psi_{\Delta_1}(t)\}] = 1 + (e^{i\xi} - 1)\int_{\Delta_1} p(t)\,dt,$$

$$E[\exp\{i\eta\psi_{\Delta_2}(t)\}] = 1 + (e^{i\eta} - 1)\int_{\Delta_2} p(t)\,dt,$$

$$E[\exp\{i(\xi\psi_{\Delta_1}(t) + \eta\psi_{\Delta_2}(t))\}] = 1 + (e^{i\xi} - 1)\int_{\Delta_1} p(t)\,dt + (e^{i\eta} - 1)\int_{\Delta_2} p(t)\,dt.$$

The last equation follows from the fact that Δ_1 and Δ_2 do not overlap. Putting $\xi = \eta = \pi$, $x = 1 - 2\int_{\Delta_1} p(t)\,dt$, $y = 1 - 2\int_{\Delta_2} p(t)\,dt$, $\varphi(x) = \Sigma q(r)x^r$ we see that (2) yields the functional equation

$$\varphi(x + y - 1) = \varphi(x)\varphi(y). \tag{3}$$

One cannot ascertain that (3) holds for all real x and y. First of all the defining power series of $\varphi(x)$ is not known to converge outside the unit circle and secondly it is not obvious that each pair of real numbers x, y between -1 and 1 is such that non-overlapping intervals Δ_1, Δ_2 exist for which

$$x = 1 - 2\int_{\Delta_1} p(t)\,dt \quad \text{and} \quad y = 1 - 2\int_{\Delta_2} p(t)\,dt.$$

[2] We use the symbol $\bar{R}$ and $E[R]$ interchangeably to denote the average (mathematical expectation) of R.

However, if one restricts oneself to small Δ_1 and Δ_2 the functional equation (3) is seen to hold in a sufficiently small neighborhood of 1. This is sufficient (in view of the analyticity of φ in the unit circle) to determine $\varphi(x)$.

In fact, differentiating (3) first with respect to x and then with respect to y we get

$$\varphi'(x)\varphi'(y) = \varphi''(x + y - 1).$$

Letting $y = 1$ and putting $\varphi'(1) = h$ we have

$$\varphi''(x) = h\varphi'(x),$$

which yields immediately

$$\varphi(x) = Ae^{hx} + B.$$

An entirely elementary reasoning (which employs the fact that $Ae^{hx} + B$ must satisfy (3)) leads to the conclusion that $B = 0$, $A = e^{-h}$ which in turn implies at once that

$$q(r) = e^{-h}\frac{h^r}{r!}.$$

Finally,

$$\int_\Delta P(t)\,dt = E[M(N;\Delta)] = E\left[\sum_{k=1}^{N} \psi_\Delta(t_k)\right]$$
$$= \left(\int_\Delta p(t)\,dt\right)e^{-h}\sum_{r=0}^{\infty}\frac{h^{r+1}}{r!} = h\int_\Delta p(t)\,dt,$$

and therefore

$$\int_{-\infty}^{+\infty} P(t)\,dt = h, \qquad P(t) = hp(t).$$

Since h is the mean value of N (i.e. $\bar{N}$) we shall use $\bar{N}$ instead of h.

3. Fourier coefficients of $F(t)$ and their statistical properties. In physical applications it is often convenient to assume that the "pulse function" $f(t)$ is periodic with period $T(T$ large) and one might therefore restrict oneself to the interval $(0, T)$.

It is furthermore assumed that both $f(t)$ and $P(t)$ are sufficiently smooth[3] so as to justify the formal operations on Fourier series performed below. Since we work in the interval $(0, T)$ we assume that $P(t) = 0$ for $t < 0$ and $t > T$.

Expanding $f(t)$ in a Fourier series in $(0, T)$ we get

$$f(t) \sim \sum_{-\infty}^{\infty} a(\omega_k)\exp(i\omega_k t), \qquad \omega_k = \frac{2\pi k}{T},$$

[3] For instance $f(t)$ and $P(t)$ may be assumed to be of bounded variation. Actually, much less severe restrictions suffice but in investigations of this sort far reaching generality would only impair the exposition.

and thus

$$F(t) \sim \sum_{-\infty}^{\infty} a(\omega_k) b(\omega_k) \exp(i\omega_k t),$$

where

$$b(\omega_k) = \sum_{j=1}^{N} \exp(-i\omega_k t_j).$$

Note that

$$E[\exp(-i\omega t)] = \int_0^T \exp(-i\omega t) p(t)\, dt$$

$$= \frac{\frac{1}{T}\int_0^T \exp(-i\omega t) P(t)\, dt}{\frac{1}{T}\int_0^T P(t)\, dt} = \frac{\rho(\omega)}{\rho(0)} = c(\omega) - is(\omega)$$

and put

$$E[F(t)] = \frac{\bar{N}}{\rho(0)} \sum_{-\infty}^{+\infty} a(\omega_k)\rho(\omega_k) \exp(i\omega_k t) = T \sum_{-\infty}^{+\infty} a(\omega_k)\rho(\omega_k) \exp(i\omega_k t),$$

$$X_k^{(\bar{N})} = \frac{\sum_{j=1}^{N} \cos(\omega_k t_j) - \bar{N} c(\omega_k)}{\sqrt{\bar{N}}},$$

$$Y_k^{(\bar{N})} = \frac{\sum_{j=1}^{N} \sin(\omega_k t_j) - \bar{N} s(\omega_k)}{\sqrt{\bar{N}}}.$$

Thus remembering that $\bar{N} = \int_0^T P(t)\, dt$ we may write

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\bar{N}}} \sim \sum_{-\infty}^{+\infty} a(\omega_k)(X_k^{(\bar{N})} - iY_k^{(\bar{N})}) \exp(i\omega_k t)$$

or

$$\frac{F(t) - \overline{F(t)}}{\sqrt{\rho(0)}} \sim \sqrt{T} \sum_{-\infty}^{\infty} a(\omega_k)(X_k^{(\bar{N})} - iY_k^{(\bar{N})}) \exp(i\omega_k t).$$

We can now state the following:

THEOREM 2. *In the limit as $\bar{N} \to \infty$ each $X_k^{(\bar{N})}$ (and $Y_k^{(\bar{N})}$) is normally distributed with mean* 0 *and variance* $\frac{1}{2} + \frac{1}{2}c(2\omega_k)$ $(\frac{1}{2} - \frac{1}{2}c(2\omega_k))$.

The proof, as usual, is based on the consideration of the characteristic function of $X_k^{(\bar{N})}$.

We have

$$E[\exp \{i\xi X_k^{(\bar N)}\}]$$

$$= \exp \{-i\xi\sqrt{\bar N}\, c(\omega_k)\} \exp(-\bar N) \sum_{r=0}^{\infty} \frac{(\bar N)^r}{r!} \left(E\left[\exp\left\{\frac{i\xi \cos \omega_k t}{\sqrt{\bar N}}\right\}\right]\right)^r$$

$$= \exp \{-i\xi\sqrt{\bar N}\, c(\omega_k)\} \exp(-\bar N) \exp\left\{\bar N E\left[\exp\left\{\frac{i\xi \cos \omega_k t}{\sqrt{\bar N}}\right\}\right]\right\}.$$

In deriving this formula use has been made of the facts that the t_j's are independent and identically distributed, that N is independent of the t_j's and that N is distributed according to Poisson's law. It is now easy to see that as $\bar N \to \infty$ the characteristic function of $X_k^{(\bar N)}$ approaches

$$\exp\{-(\tfrac{1}{4} + \tfrac{1}{4}c(2\omega_k))\xi^2\}$$

uniformly in every finite ξ-interval. This, in view of the continuity theorem for Fourier-Stieljes transforms, implies our theorem. It should be mentioned that it is tacitly assumed that even though $\bar N = T\rho(0)$ approaches ∞ it does it in such a way that the ratio $\rho(\omega)/\rho(0)$ (and hence $c(\omega)$) remains constant (or more generally, approaches a limit).

By considering the characteristic function of the joint distribution of $X_k^{(\bar N)}$ and $X_l^{(\bar N)}(|k| \neq |l|)$ (or any other pair like, for instance, $X_k^{(\bar N)}$ and $Y_l^{(\bar N)}$, in which case no restriction on k, l is necessary) we are able to prove

THEOREM 3. *In the limit as $\bar N \to \infty$ the distinct Fourier coefficients of* $(F(t) - E[F(t)])/\sqrt{\rho(0)}$ *are normally correlated (i.e. their joint distribution function is the bivariate normal distribution).*

It is also clear that the higher correlations (i.e. between more than two coefficients) will lead to multivariate normal distributions with coefficients expressible in terms of Fourier coefficients of $P(t)$.

We do not state Theorem 3 in more definite terms because in the next section we shall give a more convenient and useful way of handling correlation properties of our Fourier coefficients.

4. Statistical structure of Fourier coefficients. Let us assume that $P(t) > \gamma > 0$ and that the Fourier series of $P(t)$ converges everywhere.

Expanding $\sqrt{P(t)}$ in a Fourier series in $(0, T)$ we have

$$\sqrt{P(t)} = \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t),$$

and in particular (since $p(t) = P(t)/\bar N$)

$$\sqrt{p(t_j)} = \frac{1}{\sqrt{\bar N}} \sum_{-\infty}^{\infty} \sigma(\omega_l) \exp(i\omega_l t_j).$$

We can now write

$$b(\omega_k) = \sum_{j=1}^{N} \exp(-i\omega_k t_j) = \sum_{j=1}^{N} \frac{\exp(-i\omega_k t_j)}{\sqrt{p(t_j)}} \sqrt{p(t_j)}$$

$$= \frac{1}{\sqrt{\bar{N}}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l) \left\{ \sum_{j=1}^{N} \frac{\exp(i(\omega_l - \omega_k)t_j)}{\sqrt{p(t_j)}} \right\}$$

$$= \frac{1}{\sqrt{\bar{N}}} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} \right\}$$

$$= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k)\sigma(-\omega_l) + \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{\bar{N}}} \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - T\sigma(-\omega_l) \right\}$$

$$= T \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k)\sigma(-\omega_l)$$

$$+ \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) \left\{ \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\exp(i\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\,\sigma(-\omega_l) \right\}.$$

Put $\sigma(\omega) = \alpha(\omega) + i\beta(\omega)$, note that by Parseval's relation

$$\rho(\omega_k) = \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)\sigma(-\omega_l),$$

and introduce random variables $U_l^{(\bar{N})}$ and $V_l^{(\bar{N})}$ by means of the formulas

$$U_l^{(\bar{N})} = \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\,\alpha(\omega_l),$$

$$V_l^{(\bar{N})} = \frac{1}{\sqrt{\bar{N}T}} \sum_{j=1}^{N} \frac{\sin(\omega_l t_j)}{\sqrt{p(t_j)}} - \sqrt{T}\,\beta(\omega_l).$$

Thus

$$b(\omega_k) = T\rho(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)(U_l^{(\bar{N})} - iV_l^{(\bar{N})})l$$

and we have the following theorem.

Theorem 4. *In the limit as $\bar{N} \to \infty$ the random variables $U_0^{(\bar{N})}$, $U_1^{(\bar{N})}$, $V_1^{(\bar{N})}$, $U_2^{(\bar{N})}$, $V_2^{(\bar{N})}$, $\cdots$ are independent and normally distributed (each with mean* 0 *and variance $\frac{1}{2}$).*

This theorem can be proved in a manner exactly analogous to that of Theorem 2. We need only consider the characteristic functions of the joint distributions of U's and V's and treat them in the same way as we treated the characteristic function of the distribution of a single X in the proof of Theorem 2. One thing, however, should be strongly emphasized. The proof of independence (in the limit as $\bar{N} \to \infty$) of $U_l^{(\bar{N})}$ and $U_m^{(\bar{N})}$ ($|l| \neq |m|$), for instance, depends on proving that

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = 0.$$

This in turn depends essentially on the fact that N is distributed according to Poisson's law.

In fact,

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = \frac{1}{\bar{N}T} E\left[\left(\sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^{N} \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right] - T\alpha(\omega_l)\alpha(\omega_m).$$

But

$$E\left[\left(\sum_{j=1}^{N} \frac{\cos(\omega_l t_j)}{\sqrt{p(t_j)}}\right)\left(\sum_{j=1}^{N} \frac{\cos(\omega_m t_j)}{\sqrt{p(t_j)}}\right)\right]$$

$$= E\left[\sum_{j=1}^{N} \frac{\cos \omega_l t_j \cos \omega_m t_j}{p(t_j)}\right] + E\left[\sum_{j\neq k} \frac{\cos \omega_l t_j \cos \omega_m t_k}{\sqrt{p(t_j)}\sqrt{p(t_k)}}\right]$$

$$= E\left[\frac{N(N-1)}{(\bar{N})^2}\right] T\alpha(\omega_l)\alpha(\omega_m),$$

and finally

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = \left(\frac{\bar{N}^2 - \bar{N}}{(\bar{N})^2} - 1\right) T\alpha(\omega_l)\alpha(\omega_m).$$

Since for Poisson's distribution $\bar{N^2} = \bar{N} + (\bar{N})^2$ we get

$$E[U_l^{(\bar{N})} U_m^{(\bar{N})}] = 0.$$

Also the proof that $E[\,|\,U_l^{(\bar{N})}\,|^2] = \frac{1}{2}$ employs essentially the fact that N is distributed according to Poisson's law.

In view of Theorem 4 we can restate Theorem 3 in a form which is both useful and illuminating inasmuch as it describes completely the statistical structure of the $b(\omega_k)$'s and hence of the Fourier coefficients of $F(t)$.

THEOREM 5. *For the purposes of finding correlations between the $b(\omega_k)$'s it suffices to replace each $b(\omega_k)$ (in the limit as $\bar{N} \to \infty$) by its "statistical representation"*

$$T\rho(\omega_k) + \sqrt{T} \sum_{l=-\infty}^{\infty} \sigma(\omega_k + \omega_l)A_l,$$

where A_{-l} is the complex conjugate of A_l, A_0, A_1, A_2, $\cdots$ a sequence of independent complex-valued random variables and each A_k is distributed in such a way that $\theta_k = \arg A_k$ is uniformly distributed independent of A_k and the density of the probability distribution of $|\,A_k\,|$ is

$$2Ae^{-A^2}, \qquad (A \geq 0).$$

Theorem 5 was proved under the assumption $P(t) > \gamma > 0$. This assumption was needed to validate the convenient artifice of multiplying and dividing by $\sqrt{p(t_j)}$.

However, even in the case when $P(t)$ is not bounded from below by a positive

number (it is always true that $P(t) \geq 0$) Theorem 5 remains true. It could be proved by direct but tedious considerations suggested in section 2.

Theorems 4 and 5 can be easily extended to the case when the pulses all have the same shape but may, at random, differ in magnitude. In other words, instead of sum (1) we may consider the sum

$$F(t) = \sum_{j=1}^{N} \epsilon_j f(t - t_j), \tag{4}$$

where the individual pulses are independent and a function $P(\epsilon, t)$ is given such that

$$\int_{\epsilon}^{\epsilon+\Delta\epsilon} \int_{t}^{t+\Delta t} P(\epsilon, t)\, dt\, d\epsilon$$

is the average number of pulses of "amplitude" between ϵ and $\epsilon + \Delta\epsilon$ occurring between t and $t + \Delta t$.

Theorems 4 and 5 still hold provided one replaces the Fourier coefficients of $P(t)$ by those of

$$\int_{-\infty}^{+\infty} \epsilon P(\epsilon, t)\, d\epsilon,$$

and the Fourier coefficients of $\sqrt{P(t)}$ by those of

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{+\infty} \epsilon^2 P(\epsilon, t)\, d\epsilon}.$$

5. Concluding remarks and summary. If one assumes that the number of pulses N in the time interval $(0, T)$ is constant instead of being a random variable obeying Poisson's law, then Theorems 4 and 5 fail. The failure is due to the fact that, for instance $E[U_l^{(\bar{N})} U_m^{(\bar{N})}]$ is no longer 0. However, as $T \to \infty$ the changes in correlation due to assuming N constant become negligible. On the other hand if one assumes that the number of pulses in each of the time intervals $(0, \tau)$, $(\tau, 2\tau)$, $\cdots$ is fixed, the changes in correlations become appreciable. This case can also be treated by the above methods.

The case in which $p(t)$ is independent of time has been considered in various connections by Schottky, Uhlenbeck and Goudsmidt and Rice[4]. Their investigations emphasized the importance and usefulness of the harmonic analysis of random functions.

In conclusion we summarize our results for the case of time-dependent $P(\epsilon, t)$

[4] W. Schottky, *Ann. d. Phys.* 57 (1919) pp. 541–567.

G. E. Uhlenbeck and S. Goudsmidt, *Phys. Rev.* 34 (1929) pp. 145–151.

S. O. Rice, mimeographed notes on mathematical analysis of random noise, as yet unpublished.

The authors are indebted to Mr. Rice for making his notes available to them.

by observing that in applications one may replace $F(t)$ by its "statistical representation"

$$E[F(t)] + \sqrt{T} \sum_{k=-\infty}^{\infty} a(\omega_k) \left\{ \sum_{l=-\infty}^{\infty} \sigma(\omega_l + \omega_k) A_l \right\} \exp(i\omega_k t), \tag{5}$$

where

$$\omega_k = \frac{2\pi k}{T},$$

$$E[F(t)] = T \sum_{-\infty}^{\infty} a(\omega_k)\rho(\omega_k) \exp(i\omega_k t),$$

$$\int_{-\infty}^{\infty} \epsilon P(\epsilon, t)\, d\epsilon = \sum_{k=-\infty}^{\infty} \rho(\omega_k) \exp(i\omega_k t),$$

$$\sqrt{Q(t)} = \sqrt{\int_{-\infty}^{\infty} \epsilon^2 P(\epsilon, t)\, d\epsilon} = \sum_{k=-\infty}^{\infty} \sigma(\omega_k) \exp(i\omega_k t),$$

and the A_l's are normally distributed complex-valued random variables for which

$$E[A_l] = 0, \qquad E[|A_l|^2] = 1, \qquad A_l^* = A_{-l}.$$

Furthermore, for $l \geq 0$ the A_l's are statistically independent.

Thus

$$F(t) - E[F(t)] \sim \sum_{k=-\infty}^{\infty} \lambda(\omega_k) \exp(i\omega_k t)$$

where the λ's are normally distributed complex-valued random variables obeying the relation

$$E[|\lambda(\omega)|^2] = |a(\omega)|^2 \int_0^T Q(t)\, dt.$$

If $Q(t)$ is periodic with frequency $\dfrac{\omega_{k_0}}{2\pi}$ then it follows that $\lambda(\omega')$ and $\lambda(\omega'')$ are independent unless $\omega' + \omega''$ or $\omega' - \omega''$ is an integral multiple of ω_{k_0}.
Finally, we mention that $F(t) - E[F(t)]$ is normally distributed with variance $s(t)$ given by the formula

$$s^2(t) = E[(F(t) - E(F(t)))^2] = T \sum_{k=-\infty}^{\infty} \gamma(\omega_k)\mu(\omega_k) \exp(i\omega_k t),$$

where $\gamma(\omega_k)$ is the Fourier coefficient of $Q(t)$ and $\mu(\omega_k)$ the Fourier coefficient of $f^2(t)$.

RANDOM ALMS

By Paul R. Halmos

Syracuse University

1. Statement of the problem. Consider the problem of distributing one pound of gold dust at random among a countably infinite set of beggars. Let the beggars be enumerated and let the procedure for distribution be as follows: the first beggar is given a random portion of the gold; the second beggar gets a random portion of the remainder; $\cdots$ and so on ad infinitum. In this description the phrase "random portion" occurs an infinite number of times: it seems reasonable to require that it have the same interpretation each time. To be precise: let x_j ($j = 0, 1, 2, \cdots$) be the amount received by the jth beggar. Let the distribution of x_0 be given by a density function $p(\lambda)$:

$$p(\lambda) \geqq 0, \qquad 0 \leqq \lambda \leqq 1; \tag{1}$$

$$\int_0^1 p(\lambda)\, d\lambda = 1; \tag{2}$$

$$P(a < x_0 < b) = \int_a^b p(\lambda)\, d\lambda, \qquad 0 \leqq a < b \leqq 1. \tag{3}$$

After the first beggar has received his alms and the amount of gold dust left is μ, (i.e. $x_0 = 1 - \mu$), the value of x_1 will be between 0 and μ. The uniformity requirement mentioned above means that the proportion of μ that the second beggar is to receive is again determined by the probability density p: in other words the conditional probability that x_1 be between $\lambda\mu$ and $(\lambda + d\lambda)\mu$, given that $x_0 = 1 - \mu$, is $p(\lambda)\, d\lambda$. In symbols:

$$P(a\mu < x_1 < b\mu \mid x_0 = 1 - \mu) = \int_a^b p(\lambda)\, d\lambda. \tag{4}$$

Writing $\alpha = a\mu$, $\beta = b\mu$, (4) becomes

$$P(\alpha < x_1 < \beta \mid x_0 = 1 - \mu) = \int_\alpha^\beta \frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right) d\lambda. \tag{5}$$

More generally I shall assume that the conditional probability distribution of x_n, assuming that after the preceding donations there is left an amount μ, is given in the interval $(0, \mu)$ by $\frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right)$. In symbols:

$$P\Big(a < x_n < b \mid \sum_{j<n} x_j = 1 - \mu\Big) = \int_a^b \frac{1}{\mu}\, p\left(\frac{\lambda}{\mu}\right) d\lambda, \qquad 0 \leqq a < b \leqq \mu. \tag{6}$$

This assumption completely determines (in terms of p) the joint distribution of the whole infinite sequence $\{x_0, x_1, x_2, \cdots\}$. Several interesting special

questions may be asked about this distribution. For example: What are the expectation, dispersion, and higher moments of the x_n? What, similarly, are the moments of the partial sum $S_n = \sum_{j \leqq n} x_j$? More generally what are the exact distributions of x_n and of S_n? Will the process described really distribute all the gold, or is there a positive probability that some is left even after every beggar had his turn? What is the rate of convergence of the series $\sum_{n \geqq 0} x_n$? It is the purpose of this paper to answer these and a few related questions.

2. Calculation of distributions. The $n + 1$ dimensional probability density of the distribution of $(x_0, x_1, \cdots, x_n)$ is given by[1]

$$\prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) \tag{7}$$

in the region defined by $\lambda_j \geqq 0$, $\lambda_0 + \cdots + \lambda_n \leqq 1$. For $n = 0$ there is only one term in the product and that one is equal to $p(\lambda_0)$; the region is defined by $0 \leqq \lambda_0 \leqq 1$. The formula reduces in this case to the definition of the distribution of x_0. The general case follows inductively by the use of the conditional probability formula (6). $\left(\text{For example: } P(x_0 = \lambda_0, x_1 = \lambda_1) = P(x_0 = \lambda_0)P(x_1 = \lambda_1 \mid x_0 = \lambda_0) = p(\lambda_0) \frac{1}{1 - \lambda_0} p\left(\frac{\lambda_1}{1 - \lambda_0}\right).\right)$

From (7) it is possible in principle to calculate the densities of the distributions of x_n and of S_n. Thus for example the density q_n of the distribution of x_n is found by integrating out the λ_j with $j < n$ from (7), so that

$$q_n(\lambda_n) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) d\lambda_0 \cdots d\lambda_{n-1}, \tag{8}$$

where the integration is extended over the region defined by $\lambda_j \geqq 0$ $(0 \leqq j \leqq n)$, $\sum_{j \leqq n} \lambda_j \leqq 1$. Similarly $V_n(t) = P(S_n < t)$ is given by

$$V_n(t) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j} p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) d\lambda_0 \cdots d\lambda_n, \tag{9}$$

$(0 \leqq t \leqq 1)$ where the domain of integration is defined by $\lambda_j \geqq 0$ $(0 \leqq j \leqq n)$, $\sum_{j \leqq n} \lambda_j < t$.

Working with integrals of the type (8) and (9) is often greatly facilitated by the substitution $\mu_i = \sum_{j \leqq i} \lambda_j$, $(\lambda_i = \mu_i - \mu_{i-1})$, $0 \leqq i \leqq n$. The Jacobian of this linear change of variables is identically one. The domain of integration used in (9) is defined in terms of the μ's by $0 \leqq \mu_0 \leqq \mu_1 \leqq \cdots \leqq \mu_n \leqq t \leqq 1$, so that

$$V_n(t) = \int_0^t d\mu_n \int_0^{\mu_n} d\mu_{n-1} \cdots \int_0^{\mu_1} d\mu_0 \prod_{i \leqq n} \frac{1}{1 - \mu_{i-1}} p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right). \tag{10}$$

[1] A summation or a product extended over an empty set of indices will, as is customary, be interpreted as 0 or 1 respectively. Since throughout this paper only non-negative indices are considered, whenever the notation indicates a negative index the quantity to which it is attached is to be interpreted as 0.

Hence the density of the distribution of S_n is

$$v_n(t) = \int_0^t d\mu_{n-1} \int_0^{\mu_{n-1}} d\mu_{n-2} \cdots \int_0^{\mu_1} d\mu_0 \prod_{i<n} \frac{1}{1-\mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot \frac{1}{1-\mu_{n-1}}\, p\left(\frac{t-\mu_{n-1}}{1-\mu_{n-1}}\right). \tag{11}$$

For later purposes it is more convenient to set $t = \mu_n$ in (11) and to express $v_n(\mu_n)$ as a multiple (and not as an iterated) integral; then

$$v_n(\mu_n) = \int \cdots \int \prod_{i \leqq n} \frac{1}{1-\mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) d\mu_0 \cdots d\mu_{n-1}, \tag{12}$$

where the domain of integration is defined by $0 \leqq \mu_0 \leqq \mu_1 \leqq \cdots \leqq \mu_{n-1} \leqq \mu_n \leqq 1$. The integrals (8) and (12) are explicitly evaluated below for a special case.

It is possible from (8) to find the kth moment $M_k^{(n)}$ of x_n, $M_k^{(n)} = \int_0^1 \lambda_n^k q_n(\lambda_n)\, d\lambda_n$. Write

$$\alpha_k = \int_0^1 \lambda^k p(\lambda)\, d\lambda, \qquad \beta_k = \int_0^1 (1-\lambda)^k p(\lambda)\, d\lambda.$$

Clearly $M_k^{(n)}$ is obtained from (8) upon multiplication by λ_n^k and integration with respect to λ_n.

$$M_k^{(n)} = \int \cdots \int \lambda_n^k \prod_{i \leqq n} \frac{1}{1 - \sum_{j<i} \lambda_j}\, p\left(\frac{\lambda_i}{1 - \sum_{j<i} \lambda_j}\right) d\lambda_0 \cdots d\lambda_n. \tag{13}$$

It is advantageous once again to write $\mu_i = \sum_{j \leqq i} \lambda_j$. The resulting integral may be written in the iterated form as follows:

$$M_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i \leqq n} \frac{1}{1-\mu_{i-1}} \cdot p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot (\mu_n - \mu_{n-1})^k. \tag{14}$$

Consider separately the innermost integral

$$J = \int_{\mu_{n-1}}^1 p\left(\frac{\mu_n - \mu_{n-1}}{1-\mu_{n-1}}\right) (\mu_n - \mu_{n-1})^k \frac{d\mu_n}{1-\mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1-\mu_{n-1})$ this becomes

$$J = \int_0^1 p(\lambda)\lambda^k (1-\mu_{n-1})^k\, d\lambda = \alpha_k (1-\mu_{n-1})^k.$$

Hence

$$M_k^{(n)} = \alpha_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i<n} \frac{1}{1-\mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1-\mu_{i-1}}\right) \cdot (1-\mu_{n-1})^k. \tag{15}$$

The innermost integral this time is

$$J' = \int_{\mu_{n-2}}^{1} p\left(\frac{\mu_{n-1} - \mu_{n-2}}{1 - \mu_{n-2}}\right)(1 - \mu_{n-1})^k \frac{d\mu_{n-1}}{1 - \mu_{n-2}}.$$

Write $\lambda = (\mu_{n-1} - \mu_{n-2})/(1 - \mu_{n-2})$; then $(1 - \mu_{n-1}) = (1 - \lambda)(1 - \mu_{n-2})$ and

$$J' = \int_0^1 p(\lambda)(1 - \lambda)^k(1 - \mu_{n-2})^k\, d\lambda = \beta_k(1 - \mu_{n-2})^k.$$

Hence, finally,

$$M_k^{(n)} = \alpha_k \beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-3}}^1 d\mu_{n-2} \prod_{i<n-1} \cdot \frac{1}{1 - \mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right) \cdot (1 - \mu_{n-2})^k. \tag{16}$$

Observe now that the right member of (16) (except for the factor β_k) may be obtained from (15) upon replacing n by $n - 1$. In other words $M_k^{(n)} = \beta_k M_k^{(n-1)}$. Since $M_k^{(0)} = \alpha_k$, it follows that

$$M_j^{(n)} = \alpha_k \beta_k^n, \qquad n = 0,\ 1,\ 2, \cdots. \tag{17}$$

Instead of calculating similarly the moments $\int_0^1 \mu_n^k v_n(\mu_n)\, d\mu_n$ of S_n it is more convenient to calculate the quantities

$$N_k^{(n)} = \int_0^1 (1 - \mu_n)^k v_n(\mu_n)\, d\mu_n.$$

The moments themselves may be obtained from the N's by simple combinatorial formulas.

It follows from (12) that

$$N_k^{(n)} = \int_0^1 d\mu_0 \int_{\mu_0}^1 d\mu_1 \cdots \int_{\mu_{n-1}}^1 d\mu_n \prod_{i \leq n} \frac{1}{1 - \mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right)(1 - \mu_n)^k. \tag{18}$$

The innermost integral in (18) is

$$J'' = \int_{\mu_{n-1}}^1 p\left(\frac{\mu_n - \mu_{n-1}}{1 - \mu_{n-1}}\right)(1 - \mu_n)^k \frac{d\mu_n}{1 - \mu_{n-1}}.$$

Writing $\lambda = (\mu_n - \mu_{n-1})/(1 - \mu_{n-1})$, $(1 - \mu_n)$ becomes $(1 - \lambda)(1 - \mu_{n-1})$, so that

$$J'' = \int_0^1 p(\lambda)(1 - \lambda)^k(1 - \mu_{n-1})^k\, d\lambda = \beta_k(1 - \mu_{n-1})^k.$$

Consequently

$$\begin{aligned} N_k^{(n)} &= \beta_k \int_0^1 d\mu_0 \cdots \int_{\mu_{n-2}}^1 d\mu_{n-1} \prod_{i<n} \frac{1}{1 - \mu_{i-1}}\, p\left(\frac{\mu_i - \mu_{i-1}}{1 - \mu_{i-1}}\right) \cdot (1 - \mu_{n-1})^k \\ &= \beta_k N_k^{(n-1)}, \end{aligned} \tag{19}$$

so that

$$N_k^{(n)} = \beta_k^{n+1}, \qquad n = 0, 1, 2, \cdots. \tag{20}$$

The additivity of the first moment yields an amusing check on (17) and (20). Since $E(S_n) = E(\sum_{j\leq n} x_j) = \sum_{j\leq n} E(x_j)$ (where E denotes expectation, or first moment), it should be true that $1 - N_1^{(n)} = \sum_{j\leq n} M_1^{(j)}$. In terms of α's and β's this means $1 - \beta_1^{n+1} = \alpha_1 \sum_{j\leq n} \beta_1^j$, and this in turn reduces to the trivial identity $\alpha_1 = 1 - \beta_1$.

Since $0 \leqq \sum_{j\leq n} x_j \leqq 1$ with probability 1 for every n, it is clear that the series $\sum_{j\geq 0} x_j$ converges with probability 1 to a sum x, $0 \leqq x \leqq 1$. Since $E(x_j) = \alpha_1\beta_1^j$ and since $E(x) = \sum_{j\geq 0} E(x_j)$, it follows that $E(x) = \sum_{j\geq 0} \alpha_1\beta_1^j = \alpha_1/(1 - \beta_1) = 1$. This implies (since $0 \leqq x \leqq 1$) that x must be equal to 1 with probability 1. In other words it is almost certain that all the gold dust will eventually be distributed.

3. Product representation. Considerable light is shed on some of the above computations (and in fact the moment formulas (17) and (20) are proved anew) by the following considerations. The principle of equitable treatment enunciated in the introductory paragraph was subsequently formalized by the conditional probability relation (6). It may also be formalized by the following (equivalent) procedure. Let $y_0, y_1, y_2, \cdots$ be a sequence of *independent* chance variables each of whose distributions is given by the probability density p; let y_n be interpreted as the proportion, of the amount available to the nth beggar, that he actually receives. In other words

$$x_n = y_n\Big(1 - \sum_{j<n} x_j\Big), \qquad n = 0, 1, 2, \cdots. \tag{21}$$

The first main problem in this formulation is to express the x's in terms of the y's. This is most easily accomplished by an inductive proof of the formula

$$\sum_{j\leq n} x_j = 1 - \prod_{j\leq n} (1 - y_j). \tag{22}$$

For $n = 0$, (22) asserts merely that $x_0 = y_0$. The inductive step proceeds as follows:

$$\begin{aligned}\sum_{j\leq n} x_j &= x_n + \sum_{j\leq n-1} x_j = y_n\Big(1 - \sum_{j\leq n-1} x_j\Big) + \sum_{j\leq n-1} x_j \\ &= y_n \prod_{j\leq n-1} (1 - y_j) + 1 - \prod_{j\leq n-1} (1 - y_j) \\ &= 1 - (1 - y_n) \prod_{j\leq n-1} (1 - y_j) = 1 - \prod_{j\leq n} (1 - y_j).\end{aligned}$$

From (22) it follows that

$$x_n = y_n \prod_{j<n} (1 - y_j) \tag{23}$$

and

$$R_n = 1 - S_n = 1 - \sum_{j \leq n} x_j = \prod_{j \leq n} (1 - y_j). \tag{24}$$

The moment formulas (17) and (20) follow immediately from (23) and (24) respectively.

Another very important application of (23) and (24) is the following theorem. If the first geometric moment (geometric mean)

$$r = \exp \{E(\log [1 - y_j])\} = \exp\left\{\int_0^1 \log (1 - \lambda)p(\lambda)\, d\lambda\right\}$$

is different from zero $\left(\text{i.e. if } \int_0^1 \log (1 - \lambda)p(\lambda)\, d\lambda \text{ is finite}\right)$ then the limits

$$\lim_{n\to\infty} (x_n/y_n)^{1/n} \quad \text{and} \quad \lim_{n\to\infty} R_n^{1/n}$$

both exist and are both equal to r.

Since according to (23) and (24), $x_n/y_n = R_{n-1}$ the two parts of the conclusion are seen to be equivalent. For the proof take the logarithm of both sides of (24) and divide by n, obtaining

$$\log R_n^{1/n} = \frac{1}{n} \sum_{j \leq n} \log (1 - y_j). \tag{25}$$

Since, according to the hypotheses stated, the chance variables $\log (1 - y_j)$ are independent and all have the same distribution with a finite expectation, the strong law of large numbers applies to the right side of (25) and (after taking exponentials) yields the desired conclusion.

The result just obtained may be phrased as follows: with probability 1 x_n is asymptotically equal to $r^n y_n$. This statement shows that in an obvious if somewhat crude sense the rate of convergence of $\sum_{j \geq 0} x_j$ is that (at least) of a geometric series with ratio r. This conclusion is further supported by the behavior of R_n, which again is the sort of thing one expects from a geometric series. (That is: the nth root of the nth remainder of a geometric series always does converge to the common ratio.) As usual, more delicate quantitative results concerning the rate of convergence may be obtained by applying to (25) not merely the law of large numbers but the law of the iterated logarithm.

The product representation of x_n in formula (23) points the way to a generalization of this theory which may be of some interest. In this generalization x_n is still defined by (23) and the y's are still independent, but the distribution of y_j is given by a density p_j, where the p's need not be equal to each other. In terms of random alms this means that the condition of equitable treatment is replaced by the following weaker condition: the probability distribution of the amount that the jth beggar receives depends only on j and on the amount left by the preceding beggars, and in particular does not depend on the sizes of the alms already distributed. Many of the conclusions obtained under the simpler

hypotheses carry over to this generalized case with only slight changes. In particular the distribution formulas (7), (8), and (12), and the moment formulas (17) and (20), are changed only to the extent of acquiring an extra subscript due to the difference of the p_j.

4. Applications. (A) The original motivation of the present work was an investigation of the notion of a random mass distribution, and the results obtained may be considered as one possible solution of the problem of defining randomness for mass distributions in the special (discrete) case where the entire mass is concentrated on the non-negative integers. It would be of great interest to extend the results of this note to various continuous cases in which the set of integers is replaced by the unit interval, or the entire real line, or n dimensional Euclidean space. I intend to study some of these extensions at another time; at the moment I merely mention one implication of this statistical point of view.

Considering the sequence $\{x_0, x_1, x_2, \cdots\}$ as a system of weights, the integer n carrying the weight x_n, various questions may be raised concerning properties of the discrete mass distributions so obtained. For example: do the moments $m_k = \sum_{n \geqq 0} n^k x_n$ exist and, if so, what are their averages and dispersions and, more generally, their moments and their distributions? I shall settle here the questions concerning existence and expectation.

The chance variable m_k is non-negative and, even it if is infinite with positive probability, its expectation is defined by $E(m_k) = \sum_{n \geqq 0} n^k E(x_n) = \sum_{n \geqq 0} n^k M_1^{(n)} = \sum_{n \geqq 0} n^k \alpha_1 \beta_1^n$. Since $0 < \beta_1 < 1$, the last written series converges and therefore $E(m_k)$ is finite. This implies that m_k is finite with probability 1.

(B) It has been observed that the logarithms of the sizes of particles such as mineral grains are frequently normally distributed. Kolmogoroff[2] has given an explanation of this phenomenon; the results of the present paper yield an alternative and in some respects simpler explanation. Suppose in fact that the probability of a particle losing a chip the proportion of whose size to the size of the original particle is between λ and $\lambda + d\lambda$ is $p(\lambda)\, d\lambda$. With this stochastic scheme the size of the remaining particle after n chips have been lost is given by R_n. Since, by (25), $\log R_n$ is a sum of independent chance variables with the same distributions, the Laplace-Liapounoff theorem may be invoked to show that the distribution of R_n is for large n nearly normal. (It is necessary of course to assume here the finiteness of the second geometric moment, or equivalently of the integral $\int_0^1 \log^2 (1 - \lambda) p(\lambda)\, d\lambda$.) The mean and the variance of each summand of $\log R_n$ are

$$a = \int_0^1 \log (1 - \lambda) p(\lambda)\, d\lambda \quad \text{and} \quad b^2 = \int_0^1 [\log (1 - \lambda) - a]^2 p(\lambda)\, d\lambda,$$

[2] A. N. Kolmogoroff, "Ueber das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstueckelung," *C. R. (Doklady) Acad. Sci. URSS* (N. S.) Vol. 31(1941), pp. 99–101.

respectively; consequently (by the additivity of the mean and the variance) the corresponding parameters of the distribution of $\log R_n$ (and hence of the approximating normal distribution) are given by $(n + 1)a$ and $(n + 1)b^2$ respectively.

(C) A special case of the distributions studied in this paper (namely the case of uniform distribution, $p(\lambda) = 1$) arises in the theory of scattering of neutrons by protons of the same mass. According to Bethe[3]: "In each collision with a proton the neutron will lose energy. As long as the neutron is fast compared to the proton, the probability that the neutron energy lies between E and $E + dE$ after the collision, is $w(E)\,dE = dE/E_0$, where E_0 is the neutron energy before the collision. This means that any value of the final energy of the neutron, between 0 and the initial energy E_0, is equally probable."

To calculate explicitly the distributions it is most convenient to start from (11). If p (with any argument) is replaced by 1 and the terms of the product are distributed, each under its own differential, (11) takes the form

$$v_n(t) = \int_0^t \frac{d\mu_{n-1}}{1 - \mu_{n-1}} \int_0^{\mu_{n-1}} \frac{d\mu_{n-2}}{1 - \mu_{n-2}} \cdots \int_0^{\mu_1} \frac{d\mu_0}{1 - \mu_0}. \tag{26}$$

The value of the iterated integral is easy to obtain: $v_n(t) = (-1)^n(1/n!) \log^n (1 - t)$. Since $v_n(t)$ gives the distribution of the partial sum S_n, the distribution of $R_n = 1 - S_n$ is given by $v_n(1 - t) = (-1)^n(1/n!) \log^n t$.[4] It is possible but not necessary to derive similarly the distribution of x_n. It is simpler to obtain this distribution by exploiting the symmetry of the uniform distribution. Since, according to (23) and (24), x_n and R_n are both products of $n + 1$ uniformly and independently distributed chance variables they have the same distribution, so that the density of the distribution of x_n is also given by $(-1)^n(1/n!) \log^n t$, $n = 0, 1, 2, \cdots$.

The roles of the geometric mean r ($= 1/e$ in case $p = 1$) and of the normal distribution have also been observed in the physical situation. Fermi[5] has expressed the geometric series like behavior of $\sum_{n \geq 0} x_n$ by the statement "$\cdots$ an impact of a neutron against a proton reduces, on the average, the neutron energy by a factor $1/e$," and Bethe[6] remarks that "$\cdots$ the actual values of $\log E$ after n collisions form very nearly a Gaussian distribution $\cdots$"

---

[3] H. A. Bethe, "Nuclear Physics, B. Nuclear Dynamics, Theoretical," *Reviews of Modern Physics*, Vol. 9(1937) p. 120.

[4] This distribution has been calculated by E. U. Condon and G. Breit, "The energy distribution of neutrons slowed by elastic impacts," *Physical Review*, Vol. 49(1936) pp. 229-231.

[5] Quoted by Condon and Breit, *loc. cit.*

[6] *Loc. cit.*

ON BIASES IN ESTIMATION DUE TO THE USE OF PRELIMINARY TESTS OF SIGNIFICANCE

By T. A. Bancroft

Iowa State College

I. INTRODUCTION

In problems of statistical estimation we often express the joint frequency distribution of the sample observations $x_1, x_2, \cdots x_n$ in the form

$$f(x_1, \cdots, x_n; \alpha, \beta, \gamma, \cdots)\Pi\, dx_i, \qquad (i = 1, \cdots, n) \tag{1}$$

where the functional form, f, is assumed known, and $\alpha, \beta, \gamma, \cdots$ are certain population parameters whose values may or may not be known. Given this specification, statistical theory provides routine mathematical processes for obtaining estimates of the parameters $\alpha, \beta, \gamma, \cdots$ from the observations $x_1, x_2, \cdots, x_n$.

In performing tests of significance we often assume that the data follow some distribution

$$f_1(x_1, \cdots, x_n; \alpha, \beta, \gamma, \cdots)\Pi\, dx_i, \qquad (i = 1, \cdots, n) \tag{2}$$

where f_1 is a known function or family of functions. We may wish to test the hypothesis that the data follow the more specialized distribution

$$f_2(x_1, \cdots, x_n; \alpha', \beta', \gamma', \cdots)\Pi\, dx_i, \qquad (i = 1, \cdots, n) \tag{3}$$

where f_2 is some member or sub-group of the family f_1. Given this specification, statistical theory provides routine mathematical processes for testing such hypotheses.

In the application of statistical theory to specific data, there is often some uncertainty about the appropriate specifications in equations (1), (2) and (3). In such cases preliminary tests of significance have been used, in practice, as an aid in choosing a specification. We shall give several examples from the literature of statistical methodology.

(1) In an analysis of variance, in order to obtain a best estimate of variance, we may be uncertain as to whether two mean squares in the lines of the analysis may be assumed homogeneous, [1]. Suppose that it is desired to estimate the variance σ_1^2, of which an unbiased estimate s_1^2 is available. In addition, there is an unbiased estimate s_2^2 of σ_2^2, where from the nature of the data it is known that either $\sigma_2^2 = \sigma_1^2$ or $\sigma_2^2 < \sigma_1^2$. As a criterion in making a decision the following rule of procedure is used frequently: test s_1^2/s_2^2 by the F-test, where s_1^2 and s_2^2 are the two mean squares. If F is not significant at some assigned significance level use $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ as the estimate of σ_1^2. If F is significant at the assigned significance level, use s_1^2 as the estimate of σ_1^2.

(2) After working out the regression of y on a number of independent variates we may be uncertain as to the appropriateness of the retention of some one of

the independent variates, [2]. To illustrate let us consider the choice between the regression equations $y = b_1x_1 + b_2x_2$ and $y' = b_1'x_1$, after having fitted $y = b_1x_1 + b_2x_2$; the population regression equation being $y = \beta_1x_1 + \beta_2x_2$. In this case a procedure commonly used in deciding whether to retain x_2 is as follows: we test s_2^2/s_3^2 by the F-test, where s_2^2 is the reduction in sum of squares due to x_2 after fitting x_1, and s_3^2 is the residual mean square. If F is not significant at some assigned significance level we omit the term containing x_2 and use b_1' as the estimate of β_1. If F is significant we retain the term containing x_2 and use b_1 as the estimate of β_1. A similar example occurs in fitting a polynomial, when there is uncertainty as to the appropriate degree for the polynomial [3].

(3) In certain analyses we may be uncertain as to the appropriateness of the use of the χ^2 test. Bartlett gives an illustration in a discussion of binomial variation, [4]. He performs two supplementary χ^2 tests of significance as an aid in deciding to abandon the main use of the χ^2 test altogether, and proceeds to use an analysis of variance instead. It is of interest to note that the main use of the χ^2 test gives a significant difference at the 5% level while, in the analysis of variance, Fisher's z is not significant at the 5% level. Here again we might formulate a "rule of procedure" and follow through the analysis as in the preceding cases.

This use of tests of significance as an aid in determining an appropriate specification, and hence the form that the completed analysis shall take, involves acting as if the null hypothesis is false in those cases in which it is refuted at some assigned significance level, and, on the other hand, acting as if the null hypothesis is true in those cases in which we fail to refute it at the assigned significance level. An investigation of the consequences of some of these uses is the purpose of this paper.

It is proposed to consider the first two cases mentioned above: (1) a test of the homogeneity of variances, and (2) a test of a regression coefficient. A complete investigation of the consequences of the rules of procedure would be very extensive, since these consequences depend on the form of the subsequent statistical analysis. As a beginning, it is proposed to limit the study to the efficiency of these "rules of procedure" in the control of bias.

The need for solutions of a whole family of problems of this kind has been pointed out recently by Berkson [5].

II. EXAMPLE ONE: TEST OF HOMOGENEITY OF VARIANCES

1. Statement of the problem. s_1^2 and s_2^2 are two independent estimates of variances σ_1^2 and σ_2^2 respectively, (such that $n_1s_1^2/\sigma_1^2$, $n_2s_2^2/\sigma_2^2$ are distributed independently according to χ_1^2 and χ_2^2, with n_1 and n_2 degrees of freedom). It is known that $\sigma_2^2 \leq \sigma_1^2$. To obtain from these an estimate of σ_1^2, to be used in the particular analysis in hand, we formulate a rule of procedure.

2. Rule of procedure. Test s_1^2/s_2^2 by the F-test. If F is non-significant at some assigned significance level, we use $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ as the estimate

of σ_1^2. If F is significant at some assigned significance level we use s_1^2 as the estimate of σ_1^2. The estimate of σ_1^2 obtained by this rule of procedure will be denoted by e^*.

3. Object of this investigation. If we follow such a rule of procedure, what will be the bias in our estimate e^* of σ_1^2?

4. Derivation of the expected value of e^*. First we wish to find

$$E\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}\right), \quad \text{if} \quad \frac{s_1^2}{s_2^2} < \lambda,$$

where λ is the value on the F-distribution corresponding to some assigned significance level for n_1 and n_2 degrees of freedom.

Let $v_1 = s_1^2$, $v_2 = s_2^2$. Since s_1^2 and s_2^2 are independently distributed, the joint distribution of v_1 and v_2 is

$$c_1 v_1^{\frac{1}{2}n_1 - 1} v_2^{\frac{1}{2}n_2 - 1} \exp\left[-\frac{1}{2}\left(\frac{n_1 v_1}{\sigma_1^2} + \frac{n_2 v_2}{\sigma_2^2}\right)\right] dv_1\, dv_2,$$

where c_1 is a constant and n_1 and n_2 are the respective degrees of freedom.

Let us make the transformation of variables

$$u_1 = n_1 v_1 + n_2 v_2, \qquad 0 < u_1 < \infty$$

$$u_2 = \frac{v_1}{v_2} \qquad 0 < u_2 < \lambda,$$

then the expected value, E_1, of $\dfrac{u_1}{n_1 + n_2}$ for $u_2 < \lambda$ is given by

$$(n_1 + n_2)E_1 = \frac{c_1}{P(u_2 < \lambda)} \int_0^\lambda \int_0^\infty \frac{u_2^{\frac{1}{2}n_1 - 1}}{(n_1 u_2 + n_2)^{\frac{1}{2}(n_1 + n_2)}} \cdot u_1^{\frac{1}{2}(n_1+n_2)} \exp\left[-\frac{1}{2}\frac{u_1}{n_1 u_2 + n_2}\left(\frac{n_1 u_2}{\sigma_1^2} + \frac{n_2}{\sigma_2^2}\right)\right] du_1\, du_2$$

where $P(u_2 < \lambda)$ is the probability of u_2 being less than λ.

Integrating out u_1 and expressing the result in terms of the incomplete beta function we obtain

$$(4) \qquad (n_1 + n_2)E_1 = \frac{n_1 I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)\sigma_1^2 + n_2 I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)\sigma_2^2}{P(u_2 < \lambda)}$$

where $x_0 = (n_1\varphi\lambda)/(n_2 + n_1\varphi\lambda)$, $\varphi = \sigma_2^2/\sigma_1^2$.

We wish now to find the expected value of s_1^2 when $s_1^2/s_2^2 \geq \lambda$. Again we start with the joint distribution of v_1 and v_2, given above and this time let

$$\frac{v_2}{v_1} = Y, \qquad v_1 = v_1,$$

then the expected value, E_2, of v_1 when $Y \leqq \frac{1}{\lambda}$ is

$$E_2 = \frac{c_1}{P\left(Y \leqq \frac{1}{\lambda}\right)} \int_0^{1/\lambda} \int_0^{\infty} Y^{\frac{1}{2}n_2 - 1} v_1^{\frac{1}{2}(n_1+n_2)} \exp\left[-\frac{1}{2} v_1 \left(\frac{n_1}{\sigma_1^2} + \frac{n_2 Y}{\sigma_2^2}\right)\right] dv_1\, dY.$$

Integrating out v_1 as a gamma function, and expressing the results as incomplete beta functions we obtain

$$E_2 = \frac{[1 - I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)]\sigma_1^2}{P(Y \leq 1/\lambda)}, \tag{5}$$

where

$$x_0 = n_1\varphi\lambda/(n_2 + n_1\varphi\lambda) \text{ as before.}$$

5. Final Results. The probability that we use $(n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2)$ is $P(u_2 < \lambda)$. From equation (4) the contribution from this case to the mean value of e^* is

$$\frac{n_1 I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)\sigma_1^2 + n_2 I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)\sigma_2^2}{n_1 + n_2}.$$

The probability that we use s_1^2 is $P(Y \leq 1/\lambda)$. From equation (5) the contribution from this case is

$$[1 - I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2)]\sigma_1^2 .$$

The expected value of e^* is obtained by combining the two cases, i.e.,

$$E(e^*) = \left[1 + \frac{n_2}{n_1 + n_2}\left\{I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 1)\frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2)\right\}\right]\sigma_1^2 . \tag{6}$$

Hence the bias in e^*, expressed as a fraction of σ_1^2 is

$$\frac{n_1}{n_1 + n_2}\left[I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 1)\frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2)\right] \tag{7}$$

We note that in estimating σ_1^2 there will be a positive bias, no bias, or a negative bias according as

$$\frac{I_{x_0}(\frac{1}{2}n_1, \frac{1}{2}n_2 + 1)}{I_{x_0}(\frac{1}{2}n_1 + 1, \frac{1}{2}n_2)}$$

is greater than, equal to, or less than σ_1^2/σ_2^2.

6. Identity and checks. If $\sigma_1^2 = \sigma_2^2$, then in section 4, $E_1 = \sigma_1^2$ and

$$P(u_2 < \lambda) = I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2).$$

From (4) this gives the identity

$$(n_1 + n_2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) = n_1 I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2) + n_2 I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 1)$$

where $x_0 = n_1\lambda/(n_2 + n_1\lambda)$. This identity may be established easily by elementary calculus.

The first result in equation (6) may be checked by noting that when $\lambda = \infty$, i.e. when the two mean squares are always pooled, x_0 is 1 and equation (6) reduces to $(n_1\sigma_1^2 + n_2\sigma_2^2)/(n_1 + n_2)$. Similarly when $\lambda = 0$, in which case there is no pooling, $x_0 = 0$, and equation (6) reduces to σ_1^2.

7. Discussion. In making a choice of an appropriate estimate of σ_1^2 we may consider three procedures:

(1) Use s_1^2 always. This has the merit of having no bias, but is likely to have a large sampling error.

(2) Always pool, i.e., use $\dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$. When $\sigma_1^2 \neq \sigma_2^2$ this is biased, but in compensation will have less sampling error than (1) since it will be based on $(n_1 + n_2)$ degrees of freedom.

(3) Use the test of significance of $\dfrac{s_1^2}{s_2^2}$ as a criterion in making the decision as to whether to pool the two mean squares or not. If the test discriminates properly between cases where pooling should or should not be made, the preliminary test of significance criterion will utilize the extra n_2 degrees of freedom whenever permissible and also avoid the bias in method (2).

In Table I the expected value $E(e^*)$ divided by σ_1^2, is given for two sets of values of n_1, n_2 somewhat typical of those frequently encountered in applied work, and for a series of values of σ_2^2/σ_1^2. In addition to the case of always pooling ($\lambda = \infty$) and that of never pooling ($\lambda = 0$), the results for λ at the 1 percent, 5 percent, 20 percent levels and for $\lambda = 1$ have been tabulated. By subtracting unity from the results the bias is obtained as a fraction of σ_1^2. The Table was computed from the incomplete beta function Tables [6].

When the two mean squares are always pooled, the fractional bias is negative and increases numerically as σ_2^2 becomes small relative to σ_1^2. By examination of the values in Table I for $\sigma_2^2/\sigma_1^2 = .1$, it will be seen that the preliminary test of significance controls the bias well when σ_2^2 is much smaller than σ_1^2, that is when a large bias from pooling is most to be feared. This result happens because in such cases the preliminary test allows pooling only in a small proportion of samples.

If λ is taken at the 1 or 5 percent levels, the maximum bias appears to occur when σ_2^2/σ_1^2 is in the region 0.4–0.5, there being little bias when σ_2^2 is near σ_1^2. The lower values of λ (20 percent or λ equals 1) control the bias satisfactorily in the region $\sigma_2^2 < .6\sigma_1^2$, but have a fairly substantial positive bias when $\sigma_2^2 = \sigma_1^2$, that is when pooling would actually be justified. By use of the relation between the incomplete beta function and the sum of the terms of a binomial series it

can be shown that there is always a positive bias when $\sigma_1^2 = \sigma_2^2$ and that for given numbers of degrees of freedom this bias is greatest when $\lambda = 1$.

To summarize from the example in Table I, it seems that for small values of n_1 and n_2 none of the values of λ which have been investigated controls the bias throughout the whole range $0 \leq \sigma_2^2/\sigma_1^2 \leq 1$.

TABLE I

Expected Value of e^*/σ_1^2: $E(e^*)/\sigma_1^2$

Case 1: $n_1 = 4, n_2, = 20$

	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$	.250	.333	.417	.500	.583	.667	.750	.833	.917	1.00
$\lambda_{.01} = 4.43$	.965	.870	.791	.750	.748	.775	.821	.880	.948	1.02
$\lambda_{.05} = 2.87$	.991	.960	.924	.901	.892	.903	.930	.970	1.02	1.08
$\lambda_{.20} = 0.00$	1.00	.999	1.00	1.01	1.02	1.04	1.07	1.11	1.15	1.20
$\lambda = 1$	1.00	1.00	1.01	1.03	1.05	1.08	1.11	1.15	1.20	1.25
$\lambda = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Case 2: $n_1 = 12, n_2 = 10$

	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$	.591	.636	.682	.727	.773	.818	.864	.909	.955	1.00
$\lambda_{.01} = 4.71$	.981	.896	.833	.814	.824	.850	.884	.922	.963	1.00
$\lambda_{.05} = 2.91$	.998	.973	.935	.909	.901	.908	.928	.955	.989	1.03
$\lambda_{.20} = 0.00$	1.00	.998	.993	.987	.986	.991	1.00	1.02	1.04	1.07
$\lambda = 1$	1.00	1.00	1.00	1.00	1.01	1.02	1.04	1.06	1.08	1.11
$\lambda = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

8. The variance of e^*. Using the same method we may obtain the variance of e^*. The final result is

$$V = \frac{n_1(n_1+2)I_{x_0}(\tfrac{1}{2}n_1+2, \tfrac{1}{2}n_2)\sigma_1^4 + 2n_1n_2I_{x_0}(\tfrac{1}{2}n_1+1, \tfrac{1}{2}n_2+1)\sigma_1^2\sigma_2^2 + n_2(n_2+2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2+2)\sigma_2^4}{(n_1+n_2)^2}$$

$$(8) \qquad + \frac{n_1+2}{n_1}[1 - I_{x_0}(\tfrac{1}{2}n_1+2, \tfrac{1}{2}n_2)]\sigma_1^4 - \left[1 + \frac{n_2}{n_1+n_2}\left\{I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2+1)\frac{\sigma_2^2}{\sigma_1^2} - I_{x_0}(\tfrac{1}{2}n_1+1, \tfrac{1}{2}n_2)\right\}\right]\sigma_1^4.$$

From the relations in deriving this result the following identity was obtained:

$$(n_1 + n_2 + 2)(n_1 + n_2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2) = n_1(n_1 + 2)I_{x_0}(\tfrac{1}{2}n_1 + 2, \tfrac{1}{2}n_2) + 2n_1n_2I_{x_0}(\tfrac{1}{2}n_1 + 1, \tfrac{1}{2}n_2 + 1) + n_2(n_2 + 2)I_{x_0}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2 + 2).$$

This identity can be readily established by elementary calculus.

As a check on the result in equation (8), we note that if $\lambda = \infty$, then $x_0 = 1$, and $\frac{s_1^2}{s_2^2} < \lambda$ always. The variance of the estimate of variance becomes $2(n_1\sigma_1^4 + n_2\sigma_2^4)/(n_1 + n_2)^2$, which checks with the variance of $(n_1s_1^2 + n_2s_2^2)/(n_1 + n_2)$ for the case of always pooling. If in addition $\sigma_1^2 = \sigma_2^2$, then $V = 2\sigma_1^4/(n_1 + n_2)$. If $\lambda = 0$, then $x_0 = 0$, and $s_1^2/s_2^2 \geqq \lambda$ always. The variance of the estimate of variance becomes $2\sigma_1^4/n_1$ which checks with the variance of s_1^2 for the case of never pooling.

The expression for the variance of e^* enables us to investigate how much has been gained in terms of reduction in variance by the use of the preliminary test. The quantity $\{V + (\text{Bias})^2\}$ is the appropriate value for the whole sampling error, where V is given by (8) and the bias by (7). For the two numerical examples these quantities are shown as fractions of σ_1^4 in Table II.

As a standard of comparison the variances of the estimate s_1^2 (no pooling) will be used. In these examples the preliminary test with $\lambda = 1$ produces a variance smaller than that of s_1^2 for all values of σ_2^2/σ_1^2 except the lowest (0.1) where the two variances are equal. As λ is taken successively higher there is a substantial reduction in variance when σ_2^2 is near σ_1^2 but an increase in variance over that of s_1^2 when σ_2^2/σ_1^2 is small. Throughout nearly all the range of values of σ_2^2/σ_1^2, the smallest variance is obtained by always pooling ($\lambda = \infty$), despite the relatively large bias given by that method. This result is a reflection of the instability of estimates of variance which are based on only a few degrees of freedom.

III. EXAMPLE TWO: TEST OF A REGRESSION COEFFICIENT

1. Regression and some properties of orthogonal functions. Let

$$y = \beta_1x_1 + \beta_2x_2 + e$$

be a linear regression of y on the two variates x_1 and x_2 in which β_1 and β_2 are the respective population regression coefficients and e is the error. We assume that x_1, x_2 and y are measured from their respective sample means and that the values of x_1 and x_2 are fixed from sample to sample. In order to make comparisons among samples of different sizes we assume that x_1 and x_2 have unit variances and correlation coefficient[1] ρ so that

$$S(x_1^2) = n - 1, \qquad S(x_2^2) = n - 1, \qquad S(x_1x_2) = \rho(n - 1),$$

[1] Although ρ is commonly used to denote a population correlation coefficient, we are using it here for the sample correlation coefficient between the *fixed* variates x_1 and x_2.

TABLE II

The Variance of e^* About its True Mean: $\dfrac{V + (Bias)^2}{\sigma_1^4}$

Case 1: $n_1 = 4$, $n_2 = 20$

	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$	.577	.462	.360	.275	.205	.149	.111	.087	.076	.084
$\lambda_{.01} = 4.43$	.560	.620	.603	.523	.350	.323	.243	.184	.150	.137
$\lambda_{.05} = 2.87$	.514	.545	.554	.528	.479	.414	.353	.299	.260	.237
$\lambda_{.20} = 0.00$	.501	.500	.493	.480	.458	.435	.408	.374	.367	.360
$\lambda = 1$	.499	.493	.480	.462	.441	.423	.401	.389	.381	.387
$\lambda = 0$	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500

Case 2: $n_1 = 12$, $n_2 = 10$

	σ_2^2/σ_1^2									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
$\lambda = \infty$	.217	.183	.154	.130	.112	.097	.088	.084	.085	.091
$\lambda_{.01} = 4.71$	.185	.218	.203	.171	.141	.118	.103	.094	.092	.096
$\lambda_{.05} = 2.91$	.170	.187	.194	.183	.163	.142	.125	.114	.109	.109
$\lambda_{.20} = 0.00$	.167	.169	.171	.170	.164	.156	.146	.139	.135	.135
$\lambda = 1$	.167	.166	.165	.162	.158	.152	.148	.144	.144	.147
$\lambda = 0$	.167	.167	.167	.167	.167	.167	.167	.167	.167	.167

where $S(x_1^2)$ denotes summation of x_1^2 over the sample, with similar meanings for $S(x_2^2)$ and $S(x_1x_2)$, where n is the sample size.

We make the orthogonal transformations

$$\xi_1 = x_1, \qquad \xi_2 = x_2 - \rho x_1,$$

then

$$y = \beta_1\xi_1 + \beta_2(\xi_2 + \rho\xi_1) + e.$$

But

$$S\xi_1^2 = n - 1, \qquad S\xi_2^2 = (n-1)(1-\rho^2), \qquad S(\xi_1\xi_2) = 0,$$

therefore

$$S(y\xi_1) = \beta_1(n-1) + \beta_2\rho(n-1) + S(x_1e),$$

and

$$S(y\xi_2) = \beta_2(n-1)(1-\rho^2) + S(x_2 - \rho x_1)e.$$

Now if we represent the regression coefficients of y on the ξ's as B's we have

$$B_1 S(\xi_1^2) = S(\xi_1 y), \qquad B_2 S(\xi_2^2) = S(\xi_2 y).$$

The reduction in the total sum of squares due to x_1 ignoring x_2 is

$$B_1 S(y\xi_1) = \frac{[S(\xi_1 y)]^2}{S(\xi_1^2)} = \frac{[(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)]^2}{n-1}.$$

The reduction in the total sum of squares due to x_2 after fitting x_1 is

$$B_2 S(y\xi_2) = \frac{[S(\xi_2 y)]^2}{S(\xi_2^2)} = \frac{[(n-1)\beta_2(1-\rho^2) + S(x_2 - \rho x_1)e]^2}{(n-1)(1-\rho^2)}.$$

The reduction in the total sum of squares due to regression is

$$\frac{[(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)]^2}{n-1} + \frac{[Se(x_2 - \rho x_1) + (n-1)\beta_2(1-\rho^2)]^2}{(n-1)(1-\rho^2)},$$

in which the two parts are independently distributed.

Let b_1' be the regression coefficient of y on x_1 when the term containing x_2 is omitted from the regression equation. Now,

$$b_1' = B_1 = \frac{S(\xi_1 y)}{S(\xi_1^2)} = \frac{(n-1)(\beta_1 + \beta_2 \rho) + S(x_1 e)}{n-1}.$$

Hence

$$E(b_1') = \beta_1 + \beta_2 \rho. \tag{9}$$

Let b_2 be the regression coefficient of y on x_2 if both x_1 and x_2 are used. Then

$$b_2 = B_2 = \frac{S(\xi_2 y)}{S(\xi_2^2)} = \frac{(n-1)\beta_2(1-\rho^2) + S(x_2 - \rho x_1)e}{(n-1)(1-\rho^2)}.$$

And

$$V(b_2) = \frac{S(\xi_2^2)}{[S(\xi_2^2)]^2} = \frac{1}{S(\xi_2^2)} = \frac{1}{(n-1)(1-\rho^2)}.$$

The normal equations for $Y = b_1 x_1 + b_2 x_1$ are

$$b_1 S(x_1^2) + b_2 S(x_1 x_2) = S(x_1 y),$$

$$b_1 S(x_1 x_2) + b_2 S(x_2^2) = S(x_2 y).$$

Now

$$b_1' = \frac{S(x_1 y)}{S(x_1^2)} = b_1 + b_2 \frac{S(x_1 x_2)}{S(x_1^2)}.$$

Therefore

$$b_1' = b_1 + b_2 \rho,$$

or

$$b_1 = b_1' - b_2 \rho.$$

Therefore

$$E(b_1) = \beta_1 + \beta_2\rho - \rho E(b_2). \tag{10}$$

We notice that if $\rho = 0$, b_1 is unbiased in any selected portion of the population.

2. Statement of the problem. To obtain an estimate of b_1, in a particular analysis in hand, in which it is desirable to choose by means of a test of significance between using the regression equation $Y = b_1x_1 + b_2x_2$ and $Y' = b_1'x_1$, we formulate a rule of procedure.

3. Rule of procedure. Calculate the following analysis of variance:

	Degrees of freedom	Sum of squares	Mean square
Reduction due to x_1	1	$\dfrac{[(n-1)(\beta_1+\beta_2\rho)+S(x_1e)]^2}{n-1}$	s_1^2
Reduction due to x_2 after fitting x_1	1	$\dfrac{[(n-1)\beta_2(1-\rho^2)+S(x_2-\rho x_1)e]^2}{(n-1)(1-\rho^2)}$	s_2^2
Residual	$n-3$	$S(y-Y)^2$	s_3^2

Test $\dfrac{s_2^2}{s_3^2}$ by the F-test. If F is non-significant at some assigned significance level we omit the term containing x_2 and use

$$b_1' = \frac{(n-1)(\beta_1+\beta_2\rho)+S(x_1e)}{n-1}$$

as the estimate of β_1. If F is significant, we retain the term containing x_2, and use b_1 as the estimate of β_1. The estimate obtained by this rule will be called b^*.

4. Object of this investigation. If we follow such a rule of procedure, what will be the bias in b^* as an estimate of β_1?

5. Mathematical derivation of the bias. First, we wish $E(b_1')$ when

$$\frac{s_2^2}{s_3^2} < \lambda \quad \text{or} \quad \frac{b_2^2}{s_3^2} < \frac{\lambda}{(1-\rho^2)(n-1)},$$

where λ is the value on Snedecor's F-distribution corresponding to some assigned significance level for 1 and $(n-3)$ degrees of freedom. From (9) we have

$$E(b_1') = \beta_1 + \beta_2\rho, \tag{11}$$

no matter what the value of $\dfrac{s_2^2}{s_3^2}$; since from section 1, s_1^2 and s_2^2 are independently distributed.

Next we wish $E(b_1)$ when $\frac{s_2^2}{s_3^2} \geqq \lambda$ or $\frac{b_2^2}{s_3^2} \geqq \frac{\lambda}{(1-\rho^2)(n-1)}$. To obtain this we find it more convenient to find first $E(b_2)$ when

$$\frac{b_2^2}{s_3^2} \geq \frac{\lambda}{(1-\rho^2)(n-1)}, \quad \text{or} \quad \frac{v_3}{b_2^2} \leq \frac{1}{\lambda c_{22}},$$

where

$$v_3 = s_3^2 \quad \text{and} \quad c_{22} = \frac{1}{(n-1)(1-\rho^2)}.$$

The joint distribution of b_2, v_3 is

$$Ke^{-\frac{1}{2}(b_2-\beta_2)^2/c_{22}} v_3^{\frac{1}{2}(n-5)} e^{-\frac{1}{2}(n-3)v_3} dv_3 \, db_2,$$

where K is a constant. We make the transformation of variables

$$u = \frac{v_3}{b_2^2}, \qquad dv_3 = b_2^2 \, du;$$

then the joint distribution of b_2 and u is

$$Ke^{-(b_2-\beta_2)^2/2c_{22}} (ub_2^2)^{\frac{1}{2}(n-5)} e^{-\frac{1}{2}(n-3)b_2^2u} b_2^2 \, du \, db_2.$$

Taking the expected value when $u \leq \frac{1}{\lambda c_{22}}$ we have

$$E(b_2) = \frac{K}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)}$$

$$\cdot \int_{-\infty}^{\infty} \int_0^{1/\lambda c_{22}} b_2 \mid b_2 \mid^{n-3} u^{\frac{1}{2}(n-5)} \exp\left[-\frac{(b_2-\beta_2)^2}{2c_{22}} - \frac{(n-3)}{2} b_2^2 u\right] du \, db_2,$$

where $v_2 = s_2^2$, and $P\left(u \leq \frac{1}{\lambda c_{22}}\right)$ is the probability that u be less than or equal to $\frac{1}{\lambda c_{22}}$.

Dropping subscripts for convenience and expanding the factor which involves e to the first power of b, we have

$$E(b) = \frac{Ke^{-\beta^2/2c}}{P\left(u \leq \frac{1}{\lambda c}\right)} \int_{-\infty}^{\infty} \int_0^{1/\lambda c} b \mid b \mid^{n-3} u^{\frac{1}{2}(n-5)}$$

$$\cdot \exp\left\{-\tfrac{1}{2}b^2\left[\frac{1+(n-3)cu}{c}\right]\right\}\left[1 + \left(\frac{b\beta}{c}\right) + \frac{1}{2!}\left(\frac{b\beta}{c}\right)^2 + \cdots\right] du \, db,$$

where

$$-\infty < b < \infty, \qquad 0 < u < \frac{1}{\lambda c}.$$

Now, clearly the even terms of the series vanish whether n is odd or even when b is integrated out.

After integration with respect to b, we have an infinite series in which the typical term (apart from constants) is of the form

$$u^{\frac{1}{2}(n-5)}/[1 + (n - 3)cu]^{\frac{1}{2}(n+r)}$$

where r is an even positive integer. Subsequent integration with respect to u leads to an infinite series of incomplete integrals of the F distribution. By transforming the integrals, the series may be expressed in terms of incomplete beta functions as follows:

$$E(b_2) = \frac{\beta_2 e^{-\beta_2^2/2c_{22}}}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)} \cdot \left[I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2}\right) + \left(\frac{\beta_2^2}{2c_{22}}\right) I_{x_0}\left(\frac{n-3}{2}, \frac{5}{2}\right) + \frac{1}{2!}\left(\frac{\beta_2^2}{2c_{22}}\right)^2 I_{x_0}\left(\frac{n-3}{2}, \frac{7}{2}\right) + \cdots\right].$$

Let

$$a = \frac{\beta_2^2}{2c_{22}} \quad \text{or} \quad a = \frac{(1 - \rho^2)(n - 1)\beta_2^2}{2}.$$

Then we have

$$E(b_2) = \frac{\beta_2}{P\left(u \leq \frac{1}{\lambda c_{22}}\right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right), \tag{12}$$

where $x_0 = \dfrac{1}{\dfrac{\lambda}{n-3} + 1}$,

and λ is the desired % point of the F-distribution for 1 and $(n - 3)$ degrees of freedom. Now from (10) we have

$$E(b_1) = \beta_1 + \beta_2\rho - \rho E(b_2)$$

which enables us to obtain $E(b_1)$ from (12).

6. Final result. From (10), (11) and (12) we have

$$E(b^*) = P\left(\frac{v_2}{v_3} < \lambda\right)(\beta_1 + \beta_2\rho) + \left[1 - P\left(\frac{v_2}{v_3} < \lambda\right)\right][\beta_1 + \rho\{\beta_2 - E(b_2)\}]$$

$$= \beta_1 + \rho\beta_2 - \left[1 - P\left(\frac{v_2}{v_3} < \lambda\right)\right]\rho E(b_2).$$

The bias in b^* is

$$\rho\left[\beta_2 - \left\{1 - P\left(\frac{v_2}{v_3} < \lambda\right)\right\} E(b_2)\right].$$

Substituting the value of $E(b_2)$, we obtain

$$\text{Bias} = \rho\beta_2\left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right)\right],$$

where $x_0 = \dfrac{1}{\dfrac{\lambda}{n-3} + 1}$, $a = (1 - \rho^2)\left(\dfrac{n-1}{2}\right)\beta_2^2$.

7. Checks. From section 5 we have

$$E(b_2) = \frac{\beta_2}{P\left(\frac{v_2}{v_3} \geq \lambda\right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right),$$

where $x_0 = \dfrac{1}{\dfrac{\lambda}{n-3} + 1}$.

If $\lambda = 0$, then $x_0 = 1$, and $E(b_2) = \beta_2$.

Also from section 5 we have

$$\text{Bias} = \rho\beta_2\left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right)\right].$$

If $\lambda = 0$, then $x_0 = 1$, and Bias $= 0$.
If $\lambda = \infty$, then $x_0 = 0$, and Bias $= \rho\beta_2$.

8. Discussion. From the mathematical form of the bias,

$$\text{Bias} = \rho\beta_2\left[1 - \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right)\right],$$

where $x_0 = \dfrac{1}{\dfrac{\lambda}{n-3} + 1}$,

four deductions follow immediately: (i) There is no bias in estimating β_1, if ρ or β_2 is zero. (ii) The coefficient of β_2 in the formula is less than or at most equal to one in absolute value. (iii) The sign of the bias depends upon the signs of ρ and β_2; it is positive if both are positive or both negative, it is negative if ρ and β_2 have opposite signs. (iv) The bias is estimating β_1 is independent of β_1.

We shall discuss the bias in a few special cases by means of selected values of n, ρ, β_2 and λ. In Table III are exhibited the values of the bias for n equal to 5, 11, 21, each at ρ equal to .2, .4, .6, .8, and β_2 equal to .1, .4, 1.0, 2.0,

and 4.0. These values have been computed at the 5% point for λ, and at $\lambda = 1$. These special cases seem to indicate: (i) If we fix ρ, β_2, and λ and increase n, then the bias decreases. (ii) If we fix ρ, β_2, and n and change λ from the 5% point to $\lambda = 1$, the bias decreases considerably. (iii) If we fix ρ, n, λ and increase

TABLE III

The Bias in Estimating β_1

	$\lambda_{.05} = 18.513$, $n = 5$				$\lambda_{.05} = 5.318$, $n = 11$				$\lambda_{.05} = 4.414$, $n = 21$			
β_2 \ ρ	.2	.4	.6	.8	.2	.4	.6	.8	.2	.4	.6	.8
0.1	.017	.034	.051	.069	.015	.030	.046	.061	.014	.029	.044	.059
0.4	.067	.134	.202	.272	.049	.101	.159	.227	.033	.071	.122	.193
1.0	.142	.292	.455	.640	.028	.072	.164	.350	.001	.006	.025	.132
2.0	.162	.358	.627	1.038	.000	.000	.001	.083	.000	.000	.000	.001
4.0	.035	.101	.282	.898	.000	.000	.000	.000	.000	.000	.000	.000
	$\lambda = 1$, $n = 5$				$\lambda = 1$, $n = 11$				$\lambda = 1$, $n = 21$			
β_2 \ ρ	.2	.4	.6	.8	.2	.4	.6	.8	.2	.4	.6	.8
0.1	.004	.008	.011	.015	.004	.008	.011	.015	.004	.008	.011	.015
0.4	.012	.026	.040	.057	.009	.019	.032	.051	.005	.011	.022	.041
1.0	.011	.025	.049	.095	.001	.003	.010	.039	.000	.000	.001	.009
2.0	.000	.002	.008	.043	.000	.000	.000	.000	.000	.000	.000	.000
4.0	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

β_2 the bias increases and then decreases. This may be explained in the following manner. From section 6, the bias may be written in the form

$$\text{Bias} = \rho\beta_2 \frac{P\left(\dfrac{v_2}{v_3} < \lambda\right)}{P\left(\dfrac{v_2}{v_3} > \lambda\right)} \sum_{i=0}^{\infty} \frac{a^i e^{-a}}{i!} I_{x_0}\left(\frac{n-3}{2}, \frac{3}{2} + i\right).$$

Now if ρ, n, λ are held constant and β_2 is relatively small, $P\left(\dfrac{v_2}{v_3} < \lambda\right)$ is relatively large and $\sum_{i=0}^{\infty} \dfrac{a^i e^{-a}}{i!} I_{x_0}\left(\dfrac{n-3}{2}, \dfrac{3}{2} + i\right)$ is relatively large, but $P\left(\dfrac{v_2}{v_3} > \lambda\right)$ is relatively small. Hence, for a while as we increase β_2 the bias will increase, but as β_2 gets larger $P\left(\dfrac{v_2}{v_3} < \lambda\right)$ and $\sum_{i=0}^{\infty} a^i e^{-a} I_{x_0}\left(\dfrac{n-3}{2}, \dfrac{3}{2} + i\right)$ becomes smaller while $P\left(\dfrac{v_2}{v_3} > \lambda\right)$ becomes larger. Hence, a value of β_2 will be reached at which the

bias begins to decrease. (iv) If we fix n, β_2, and λ and increase ρ, the bias increases without exception.

The above results were obtained under the assumption that a test of significance criterion is used in making a choice as to the number of independent variables to be retained after the regression $y = b_1x_1 + b_2x_2$ has been fitted. If this test of significance criterion is used, we may wish to have a means of controlling the bias. From a study of Table III we note that the bias may be decreased by increasing n and by using $\lambda = 1$. We also notice that as β_2 increases from 0.1 to 4.0 the bias increases and then decreases; and so passes through a maximum value. Hence, if we have a regression in which β_2 is fairly well below or above this maximum value, we would expect a smaller bias.

The bias in estimating β_1 is "unstudentized," i.e., is a function of the population parameters ρ and β_2. In any particular analysis in hand, it would be necessary to know the values of ρ and β_2 or be willing to use estimates of them obtained from the data.

It is realized that only a beginning has been made on the regression problem: an investigation should be undertaken of the more general problem of the use of a test of significance criterion in making a choice as to the number of independent variables to be retained after the regression

$$y = b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

has been fitted.

REFERENCES

[1] J. Wishart and A. R. Clapham, "A study in sampling techniques: the effect of artificial fertilizers on the yield of potatoes," *Jour. Agri. Sci.*, Vol. 19 (1929), Part 4, p. 605.

[2] W. G. Cochran, "The omission or addition of an independent variate in multiple linear regression," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 5 (1938), p. 171.

[3] G. W. Snedecor, *Statistical Methods*, Third edition. Iowa State College Press, Ames, Iowa. 1940 Sec. 14.3.

[4] M. S. Bartlett, "Square root transformation in analysis of variance," *Jour. Roy. Stat. Soc. Suppl.*, Vol. 3 (1936), pp. 76–77.

[5] Joseph Berkson, "Tests of significance considered as evidence," *Jour. Amer. Stat. Assn.*, Vol. 37 (1942), pp. 325–335.

[6] Karl Pearson (Editor), *Tables of the Incomplete Beta Function*, Cambridge Univ. Press. *Biometrika*. Univ. Coll., London, 1934.

THE PROBABILITY OF CONVERGENCE OF AN ITERATIVE PROCESS OF INVERTING A MATRIX

By Joseph Ullman

Columbia University

Introduction. The inversion of a matrix is a computational problem of wide application. This is a further study of an efficient iterative method of matrix inversion described by Harold Hotelling [1], with an examination of the probability of convergence in relation to the accuracy of the initial approximation. The lines of investigation were suggested both by his article and by helpful comments made during the course of the research.

The inverse of a matrix can be obtained to any desired degree of accuracy by using a variation of the Doolittle method, and starting with a sufficient number of accurate decimal places in the matrix being inverted. This procedure becomes inefficient if the order of the matrix is large, or if the desired degree of accuracy is very great. In either case the efficiency can be greatly increased by first obtaining an approximation to a small number of decimal places and then applying a method of iteration until the desired accuracy is achieved.

1. Iterative methods. Hotelling's method of iteration is as follows. Let A be the matrix to be inverted and let C_0 be the approximation to the inverse. Calculate in turn $C_1, C_2, \cdots$ where,

$$C_{m+1} = C_m(2 - AC_m). \tag{1.1}$$

This sequence of matrices will converge to the inverse of A if the roots of

$$D = 1 - AC_0, \tag{1.2}$$

are all less than unity in absolute value.

The iterative method (1.1) will be generalized to yield a class of iterative methods, one element of which will be shown to be more efficient, in certain cases, than method (1.1). The generalized iterative method is,

$$C_{m+1} = C_m\{1 + (1 - AC_m) + (1 - AC_m)^2 + \cdots + (1 - AC_m)^{k-1}\}. \tag{1.3}$$

For every k, the condition for convergence is that the roots of the matrix (1.2) all be less than unity in absolute value.

A method of comparing the efficiency of these different iterative methods arises from the following considerations. Since

$$C_0 = A^{-1}(AC_0), \tag{1.4}$$

which is equivalent to

$$C_0 = A^{-1}(1 - D), \tag{1.5}$$

it follows that

$$A^{-1} = C_0(1 - D)^{-1}. \tag{1.6}$$

When the roots of D are all less than unity in absolute value, (1.6) has the infinite expansion,

$$A^{-1} = C_0(1 + D + D^2 + D^3 + \cdots). \tag{1.7}$$

The general iterative process (1.3) generates the infinite series in the following manner,

$$(1 + D + D^2 + \cdots + D^{k-1})(1 + D^k + \cdots + D^{k(k-1)}) \cdot (1 + D^{k^2} + \cdots + D^{k^2(k-1)}) \cdots . \tag{1.8}$$

Each parentheses corresponds to one iteration. Hence k^m terms are generated by m iterations. In order to achieve the accuracy of n terms in (1.7), $m = \log_e n/\log_e k$ iterations are required. Each iteration involves k matrix multiplications, so that $km = k \log_e n/\log_e k$ is the total number of matrix multiplications necessary to achieve this degree of accuracy.

The integer for which this is a minimum is three. Therefore the "most efficient" method of iteration is,

$$C_{m+1} = C_m\{1 + (1 - AC_m) + (1 - AC_m)^2\}. \tag{1.10}$$

If the desired degree of accuracy can be achieved by one application of (1.1), or by two applications of (1.1) but not by one application of (1.10), then (1.1) is preferable.

2. The condition for convergence. The sequence,

$$C_1, \quad C_2, \quad C_3, \quad \cdots \tag{2.1}$$

obtained from (1.3) will converge to the inverse of A if the roots of

$$D = 1 - AC_0, \tag{2.2}$$

are all less than unity in absolute value. The following assumptions determine the nature of D.

We assume that the expected value of each element of the first approximation C_0 is equal to the corresponding element of the exact inverse of A. The actual values of the elements of C_0 will deviate from their expected values. We will consider two important cases. If the deviations are entirely due to the fact that the elements of C_0 are only accurate to a limited number of decimal places, say k, then the deviations may be regarded as distributed with constant density over a range of length 10^{-k}. It will be assumed that the deviations of the elements of C_0 from their expected values are independent. While this case arises in practice, we will first treat a closely related case, which lends itself to exact treatment more readily. We assume that the deviations of the elements of C_0

are normally distributed about their expected values, with variance $\mu = 10^{-2k}/12$. The variance μ is the same as that which arises if the probability density is uniform with range 10^{-k}.

The elements of E, the matrix of deviations,

$$E = A^{-1} - C_0, \tag{2.3}$$

are independently and normally distributed. Combining (2.2) and (2.3) we obtain

$$D = 1 - AC_0 = A(A^{-1} - C_0) = AE. \tag{2.4}$$

Let p be the order of the matrix A. Each element of D will be a linear combination of p independently and normally distributed variables, and therefore will itself be normally distributed. A sufficient condition for all the roots of D to be less than unity in absolute value, and hence for the process of iteration to converge, is for the sum of the squares of the elements of D to be less than unity in absolute value. We will use the following notation

(2.5) d_{ij}: the element of D in the ith row and jth column,

$$N_D^2 = \sum_i \sum_j d_{ij}^2 .$$

A procedure suggested by this relationship is to determine the probability distribution of N_D^2, so that probability statements concerning the absolute value of the roots of D can be made. Because the elements of D are not all independent, no multiple of N_D^2 can be expected to have the $\chi^2_{(p^2)}$ distribution.[1]

The distribution of N_D^2 is shown to be closely related to the chi-square distribution in the next section, and on the basis of this relationship, lower bounds to the probability of convergence of the iterative process are developed in section 4. In section 5 the exact distribution of the norm is obtained for a general class of cases. The final section is concerned with the validity of applying the results of this study to a practical situation, where the deviations of the elements of C_0 from their expected values are uniformly, rather than normally, distributed.

3. An equivalance. Let e_{ij} be the element of E in the ith row and the jth column, and a_{ij} be the element of A in the ith row and the jth column. From (2.4) and (2.5), we find that

$$d_{ij} = \sum_k a_{ik} e_{kj} .$$

Since the elements of E are independently and normally distributed with variance $\mu = 10^{-2k}/12$ it follows readily that

$$E[e_{ij}e_{kh}] = \delta_{ik}\delta_{jh}\mu. \tag{3.2}$$

[1] The number in the parentheses will indicate the number of degrees of freedom of the chi-square distribution.

Making use of (3.1) and (3.2), we find that for two d_{ij} in the same column,

$$E[d_{ij}d_{kj}] = \mu \sum_t a_{it}a_{kt}, \tag{3.3}$$

while for any two d_{ij} in different columns,

$$E[d_{ij}d_{kh}] = 0. \tag{3.4}$$

From (3.3) and (3.4) it follows that it is permissible to regard the elements of the p columns of D as the coordinates of p independently selected points from a multivariate normal universe with covariance matrix $\sigma = \mu AA'$. We will let $\lambda = \sigma^{-1}$.

The moment generating function of the sum of squares of the coordinates of any point is

$$\frac{|\lambda|^{\frac{1}{2}}}{|\lambda - 2t|^{\frac{1}{2}}}. \tag{3.5}$$

This can also be written as

$$\frac{1}{(1 - 2\sigma_1 t)^{\frac{1}{2}}(1 - 2\sigma_2 t)^{\frac{1}{2}} \cdots (1 - 2\sigma_p t)^{\frac{1}{2}}}, \tag{3.6}$$

where $\sigma_1, \cdots, \sigma_p$ are the characteristic roots of σ.

Since N_D^2 is the sum of p independent expressions of this type, its moment generating function is the pth power of (3.6),

$$\frac{1}{(1 - 2\sigma_1 t)^{\frac{1}{2}p} \cdots (1 - 2\sigma_p t)^{\frac{1}{2}p}}. \tag{3.7}$$

This expression is the moment generating function of

$$\sigma_1 \chi^2_{(p)1} + \sigma_2 \chi^2_{(p)2} + \cdots + \sigma_p \chi^2_{(p)p}, \tag{3.8}$$

where the $\chi^2_{(p)i}$ are all independent.

Writing the roots as

$$\sigma_0, \sigma_0 - k_1, \cdots, \sigma_0 - k_{p-1}, \tag{3.9}$$

where σ_0 is the largest root of σ, and all $k_i > 0$, it follows that N_D^2 has the same distribution as

$$\sigma_0 \sum_i \chi^2_{(p)i} - \sum_{j=1}^{p-1} k_j \chi^2_{(p)j}. \tag{3.10}$$

Therefore, making use of the reproductive power of χ^2, we obtain

$$\begin{aligned} P\{N_0 < 1\} &= P\left\{\sigma_0 \sum_i \chi^2_{(p)i} < 1 + \sum_{j=1}^{p-1} k_j \chi^2_{(p)j}\right. \\ &= P\left\{\sigma_0 \chi^2_{(p^2)} < 1 + \sum_{j=1}^{p-1} k_j \chi^2_{(p)j}\right\}. \end{aligned} \tag{3.11}$$

By making special assumptions about the k_i, close approximations to the probability that N_D will be less than one, and hence that the process of iteration will converge, can be obtained. Instead of following this procedure, it is more desirable to have definite lower bounds for the probability that N_D will be less than one. This will lead to an overstatement in the number of decimal places of accuracy necessary in the first approximation C_0 to assure convergence, but it will practically eliminate the possibility of having to recalculate the first approximation, and hence will lead to greater efficiency in the long run.

4. The derivation of the formula for determining the required degree of accuracy. The inequality used in this section is derived in two steps from (3.10). Since $k_i > 0$ $(i = 1, \cdots, p - 1)$ it follows immediately that

$$P\{N_D < 1\} > P\{\sigma_0\chi^2_{(p^2)} < 1\}. \tag{4.1}$$

In order to use this inequality, the upper bound for σ_0

$$\sigma_0 \leq (\mathrm{tr}\ \sigma^t)^{1/t} \tag{4.2}$$

can be used. For $t = 1$,

$$\sigma_0 \leq \mathrm{tr}\ \sigma = \mathrm{tr}\ \mu AA' = \mu\ \mathrm{tr}\ AA' = \mu N_A^2 . \tag{4.3}$$

Dr. Wald pointed out that using (4.2) for $t = 1$ reduces the amount of information retained in (4.1) to that which is contained in the inequality,

$$N(D) \leq N(A)N(E). \tag{4.4}$$

A closer upper bound is feasible in any particular case, and can be introduced at this point by letting $t = 2$ or $t = 3$. The following formula will be developed for the general case, making use of (4.3).
Substituting (4.3) in (4.1), we obtain

$$P\{N_D < 1\} > P\left\{\chi^2_{(p^2)} < \frac{1}{\mu N_A^2}\right\}. \tag{4.5}$$

It is desirable to separate the effects of the order of the matrix A on convergence, and the order of magnitude of the elements. Hence we introduce as a measure of the average size of the a_{ij} their root mean square m, so that

$$m^2 = \sum_i \sum_j a_{ij}^2/p^2. \tag{4.6}$$

Hence

$$N_A = pm. \tag{4.7}$$

The final form of the inequality is

$$P\{N_D < 1\} < P\left\{\chi^2_{(p^2)} < \frac{12\cdot 10^{2k}}{p^2m^2}\right\}. \tag{4.8}$$

First we will obtain an expression for the number of decimal places required in the first approximation to make the probability of convergence at least .999. Then the expression will be checked directly by means of (4.8) and tables of the chi-square function.

For large values of p, $\sqrt{2\chi^2_{(p^2)}}$ is approximately normally distributed with mean value $\sqrt{2p^2 - 1}$ and unit variance [2], [5]. Applying this transformation to the right hand side of (4.8), and noting that 3.1 standard deviations is slightly greater than the deviation corresponding to .999, we obtain as the condition for

$$P\left\{\chi^2_{(p^2)} < \frac{12\cdot 10^{2k}}{p^2m^2}\right\} \geq .999 \tag{4.9}$$

or

$$P\left\{2\chi^2_{(p^2)} < 2\cdot\frac{12\cdot 10^{2k}}{p^2m^2}\right\} \geq .999 \tag{4.10}$$

that it is sufficient that

$$\sqrt{\frac{24\cdot 10^{2k}}{p^2m^2}} - \sqrt{2p^2 - 1} \geq 3.1. \tag{4.11}$$

This is equivalent to

$$k \geq \log_{10} p + \log_{10} m + \log_{10}\left(\sqrt{p^2 - \tfrac{1}{2}} + \frac{3.1}{\sqrt{2}}\right) - \tfrac{1}{2}\log_{10} 24 + \log_{10}\sqrt{2}. \tag{4.12}$$

Since the characteristic of a logarithm is insensitive to the argument, rounding off will introduce a negligible error, and we finally obtain an upper limit to the lower bound of k,

$$k > \log_{10} m + \log_{10} p + \log_{10} (p + 3) - .55. \tag{4.13}$$

In order to verify the accuracy of (4.13) for small values of p, certain values of p, k and m are chosen and the probabilities associated with (4.8) determined [2]. The entries in brackets are the corresponding values of k determined from (4.13).

A typical example will illustrate the use of table on facing page. Let the matrix A to be inverted be a fourth order correlation matrix. The mean magnitude m is about $\frac{1}{2}$ and $p = 4$. If the first approximation C_0 is obtained to one place accuracy, then the probability that the sequence $C_1, C_2, \cdots$ will converge to A^{-1} will be greater than .999. Using formula (4.13), we obtain $k = .53$. Since one is the first integer greater than .53, the table verifies the use of the formula.

Although the formula was developed on the assumption that p is large, every value calculated is consistent with the table. This lends support to its use for small values of p.

*The Probability of Convergence of the Iterative Process**

	k \ p	2	3	4	5
	-1	0+	0+	0+	0+
$m = \frac{1}{2}$	0	[.05].982	[.33].199	[.53]0+	[.70]0+
	1	1−	1−	1−	1−
	-1	0+	0+	0+	0+
$m = 2$	0	[.85].051	[.93]0+	0+	0+
	1	1−	1−	[1.13].715	[1.30]0−
	2	1−	1−	1−	1−
	0	0+	0+	0+	0+
$m = 10$	1	[1.35].439	[1.63]0+	1[.83]0+	0+
	2	1−	1−	1−	[2.00]1−

* "1−" means greater than .999.

It has already been pointed out that k is not sensitive to rounding off of the argument of the logarithm. Thus for $p = 20$ and $m = 2$, we can let $\log_{10} m = .3$, $\log_{10} p = 1.3$, $\log_{10} (p + 3) = 1.36$ and obtain

$$k = .3 + 1.3 + 1.36 - .55 = 2.41,$$

from which it follows that three decimal place accuracy in C_0 will practically insure convergence of the iterative process.

5. The mean, variance, and exact distribution. To obtain the moments of N_D^2, the most convenient form to use is (3.8). Since the $\chi^2_{(p)i}$ are independent

$$E[N_0^2] = E[\sum_i \sigma_i \chi^2_{(p)i}] = p \sum_i \sigma_i . \tag{5.1}$$

$$\begin{aligned} \sigma_{N_D^2} &= E[N_D^2] - (E[N_D^2])^2 \\ &= E\left[\sum_{i=1}^{p} \sigma_i^2(\chi^2_{(p)i})^2 + 2\sum_{i<j} \chi^2_{(p)i}\chi^2_{(p)j}\sigma_i\sigma_j\right] - (p\sum_i \sigma_i)^2 \\ &= (2p + p^2)\sum_i \sigma_i^2 + 2p^2 \sum_{i<j} \sigma_i\sigma_j - p^2 \sum_i \sigma_i^2 - 2p^2 \sum_{i<j} \sigma_i\sigma_j \\ &= 2p \sum_i \sigma_i^2 . \end{aligned} \tag{5.2}$$

These can be expressed in terms of the elements of A and the variance of the elements of E, since

$$\begin{aligned} \sum_i \sigma_i &= \operatorname{tr}(\sigma) = \mu \operatorname{tr}(AA') = \mu N_A^2 , \\ \sum_i \sigma_i^2 &= \operatorname{tr} \sigma^2 = \mu^2 \operatorname{tr}(AA'AA'). \end{aligned} \tag{5.3}$$

The exact distribution of N_D^2 can be obtained readily when p is even. In this case the infinite integral,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-i\,t\,N_D^2}\,dt}{(1-i2\sigma_1 t)^{p/2}\cdots(1-i2\sigma_p t)^{p/2}}, \tag{5.4}$$

can be evaluated by contour integration. The integral satisfies the conditions given in Whittaker and Watson [3, sec. 622], if the semicircle of the contour is taken on the lower half of the complex plane.

For the case $p = 2$, for example, there are simple poles, at $t = \dfrac{-i}{2\sigma_1}, \dfrac{-i}{2\sigma_2}$. The sum of the residues at these poles, multiplied by i yields the exact distribution:

$$\frac{\sigma_1 e^{-N_D^2/2\sigma_1}}{2(\sigma_1-\sigma_2)}+\frac{\sigma_2 e^{-N_D^2/2\sigma_2}}{2(\sigma_2-\sigma_1)}. \tag{5.5}$$

For even values of p greater than 2, the values of the residues can be obtained by repeated differentiation.

6. Summary. We are now in a position to discuss the applicability of the results of this paper to the problem which arises most frequently in practice. The elements of the first approximation to the inverse will deviate from their expected values only because the first approximation is carried to a limited number of places, say k. In this case the deviations will be distributed with constant density over a range of length 10^{-k}. The elements of E, the matrix of deviations,

$$E = A^{-1} - C_0, \tag{6.1}$$

are now each independently distributed, but with uniform density, range 10^{-k} and mean equal to zero. From (2.4)

$$D = AE, \tag{6.2}$$

we observe that each element of D will be a linear combination of p independently and rectangularly distributed variables, each with mean zero and range 10^{-k}. The analysis of sections 3, 4, and 5 will be valid if d_{ij} can be considered to be normally distributed.

There is much experimental evidence and theoretical justification for assuming that the elements of D are normally distributed. A sufficient condition that the d_{ij} approach normality as p increases is that the sum of the a_{ij}^2 in any row of A be divergent as the order of the matrix approach infinity, while at the same time every element be less than some constant value independent of the order of the matrix [4].

The experimental and theoretical evidence supporting the approach of the d_{ij} to normality, the fact that the logarithms are insensitive to errors of approxi-

mation in their arguments and the fact that the lower bounds to the probability of convergence of the iterative process are used, all lend support to the formula

$$k > \log_{10} m + \log_{10} p + \log_{10} (p + 3) - .55.$$

for determining the number of places (k) necessary in the first approximation (C_0) to the inverse of A, a matrix of order p whose elements have mean size m, to make the probability at least .999 that the process of interation will yield a sequence of matrices which will converge to the true inverse. The ultimate justification of the use of this formula can only be by the results of its application in practice.

REFERENCES

[1] HAROLD HOTELLING, "Some new methods in matrix calculation," *Annals of Math. Stat.*, Vol. 14 (1943) pp. 1–34.
[2] FISHER and YATES, *Statistical Table for Biological, Agricultural and Medical Research*, p. 31.
[3] WHITTAKER and WATSON, *Modern Analysis*, 4th edition, pp. 113–114.
[4] J. V. USPENSKY, *Introduction to Mathematical Probability*, 1935, Chapter 14.
[5] E. B. WILSON and M. M. HILFERTY, "The distribution of chi-square," *Proc. Nat. Acad. Science*, Vol. 17 (1931), pp. 684–688.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

ON DISTRIBUTION-FREE TOLERANCE LIMITS IN RANDOM SAMPLING

By Herbert Robbins

Post Graduate School, Annapolis, Md.

Let $X_1, \cdots, X_n$ be independent random variables each with the continuous and differentiable cumulative distribution function $\sigma(x) = Pr(X_i < x)$. A continuous function $f(x_1, \cdots, x_n)$ with the property that the random variable $Y = \sigma(f(X_1, \cdots, X_n))$ has a probability distribution which is independent of $\sigma(x)$ will be called a *distribution-free upper tolerance limit*[1] (d. f. u. t. l.). We shall prove

Theorem 1. *A necessary and sufficient condition that the continuous function $f(x_1, \cdots, x_n)$ be a d. f. u. t. l. is that the function*

$$\bar{f}(x_1, \cdots, x_n) = \prod_{i=1}^{n} \{f(x_1, \cdots, x_n) - x_i\}$$

be identically zero.

Proof. Since f is continuous, we can prove the necessity of the condition by deriving a contradiction from the assumption that f is a d. f. u. t. l. for which there exist distinct numbers $a_1, \cdots, a_n$ such that $f(a_1, \cdots, a_n) = A \neq a_i$, $(i = 1, \cdots, n)$.

Since the numbers $a_1, \cdots, a_n, A$ are distinct, there will exist a positive number ϵ such that the $(n + 1)$ intervals

$$I: \quad A - \epsilon \leq x \leq A + \epsilon$$

$$I_i: \quad a_i - \epsilon \leq x \leq a_i + \epsilon \qquad (i = 1, \cdots, n),$$

have no points in common. Moreover, since f is continuous, there will correspond to ϵ a positive number $\epsilon_1 < \epsilon$ such that

$$A - \epsilon \leq f(x_1, \cdots, x_n) \leq A + \epsilon,$$

provided that simultaneously

$$|x_i - a_i| < \epsilon_1 \qquad (i = 1, \cdots, n).$$

Now let p be any number between $\frac{1}{3}$ and $\frac{2}{3}$. Corresponding to p we define the function $\sigma_p(x)$ as follows. In the interval I we set $\sigma_p(x) = p$. In every interval

$$J_i: \quad a_i - \epsilon_1 \leq x \leq a_i + \epsilon_1 \qquad (i = 1, \cdots, n)$$

[1] Cf. S. S. Wilks, *Mathematical Statistics*, Princeton University Press (1943), pp. 93–94.

we let $\sigma_p(x)$ increase an amount $\left(\frac{1}{3n}\right)$. Outside the intervals $I, J_1, \cdots, J_n$ we define $\sigma_p(x)$ in any manner so that it is continuous, differentiable, and non-decreasing for every x, and has the properties $\sigma_p(-\infty) = 0$, $\sigma_p(\infty) = 1$. It is clear that we can do this.

Let S denote the set of all points $(x_1, \cdots, x_n)$ of n dimensional space such that simultaneously

$$|x_i - a_i| \leq \epsilon_1 \qquad (i = 1, \cdots, n).$$

Then by construction, for $\sigma_p(x)$ defined above,

$$Pr((X_1, \cdots, X_n) \epsilon S) = \left(\frac{1}{3n}\right)^n.$$

But if $(X_1, \cdots, X_n) \epsilon S$, then by construction,

$$A - \epsilon \leq f(X_1, \cdots, X_n) \leq A + \epsilon$$

and

$$Y = \sigma_p(f(X_1, \cdots, X_n)) = p.$$

Hence for $\sigma(x) = \sigma_p(x)$ we have

$$Pr(Y = p) \geq \left(\frac{1}{3n}\right)^n.$$

But since f is a d. f. u. t. l., this inequality must hold for any $\sigma(x)$.

Now choose a set of numbers

$$\tfrac{1}{3} < p_1 < p_2 < \cdots < p_m < \tfrac{2}{3},$$

where $m = 2(3n)^n$. Then from the above,

$$Pr(Y = \text{one of the numbers } p_1, \cdots, p_m) \geq 2.$$

This is the desired contradiction.

Let $O_r(x_1, \cdots, x_n)$ be the function whose value is the rth term when the numbers $x_1, \cdots, x_n$ are arranged in non-decreasing order of magnitude. In terms of the functions O_r we can characterize the continuous functions f which satisfy the identity $\tilde{f} \equiv 0$ as follows. Let $i_1, \cdots, i_n$ be a permutation of the integers $1, \cdots, n$. Denote by $E(i_1, \cdots, i_n)$ the set of all points $(x_1, \cdots, x_n)$ such that

$$x_{i_1} < x_{i_2} < \cdots < x_{i_n}.$$

The $n!$ sets E are open and disjoint. Since f is continuous and $\tilde{f} \equiv 0$, in each $E(i_1, \cdots, i_n)$ we must have, for some r,

$$f(x_1, \cdots, x_n) \equiv O_r(x_1, \cdots, x_n),$$

where the integer $r = r(i_1, \cdots, i_n)$ must depend on the permutation $i_1, \cdots, i_n$ in such a way that f may be extended continuously over the whole space. (The

condition for this is as follows. Two permutations $i_1, \cdots, i_n$ and $j_1, \cdots, j_n$ may be called *adjacent* if they differ only by an interchange of two adjacent integers. Then for any two adjacent permutations, either $r(i_1, \cdots, i_n) = r(j_1, \cdots, j_n)$ or the two values of r are the two interchanged integers. For example, the function

$$f(x_1, x_2, x_3) = \begin{cases} O_3(x_1, x_2, x_3) & \text{if} \quad O_3(x_1, x_2, x_3) = x_1 \\ O_2(x_1, x_2, x_3) & \text{otherwise} \end{cases}$$

satisfies this requirement.)

We shall now prove that the necessary condition, $\bar{f} \equiv 0$, of Theorem 1 is sufficient to ensure that the continuous function f be a d. f. u. t. l. From the argument of the preceding paragraph, any continuous function f such that $\bar{f} \equiv 0$ will in each set $E(i_1, \cdots, i_n)$ have the value $O_r(x_1, \cdots, x_n)$, where r is an integer from 1 to n. Since the variables $X_1, \cdots, X_n$ are independent and have the same probability distribution, the probability that $(X_1, \cdots, X_n)$ will belong to $E(i_1, \cdots, i_n)$ is equal to $(1/n!)$ for every permutation $i_1, \cdots, i_n$. Let

$$W = f(X_1, \cdots, X_n).$$

Then if $\varphi(x) = d\sigma(x)/dx$ denotes the probability density function of each X_i, the conditional p. d. f. of $W = O_r(X_1, \cdots, X_n)$, given that $(X_1, \cdots, X_n)$ belongs to $E(i_1, \cdots, i_n)$, will be $n!\psi_r(w)$, where

$$\psi_r(w) = \frac{\varphi(w)\sigma^{r-1}(w)[1 - \sigma(w)]^{n-r}}{(r-1)!\,(n-r)!}.$$

Thus $\psi_r(w)$ will be of the form

$$\psi_r(w) = \varphi(w)F_r(\sigma(w)),$$

where $F_r(\sigma(w))$ is a polynomial in $\sigma(w)$. Hence the conditional p. d. f. of $Y = \sigma(W)$, given that $(X_1, \cdots, X_n)$ belongs to $E(i_1, \cdots, i_n)$, will be $n!\xi_r(y)$, where

$$\xi_r(y) = F_r(y),$$

and the p. d. f. of Y will be

$$\xi(y) = \Sigma F_r(y),$$

where the summation is over the $n!$ integers $r = r(i_1, \cdots, i_n)$. This is independent of $\sigma(x)$, so that f is a d. f. u. t. l. This completes the proof of Theorem 1.

A function $f(x_1, \cdots, x_n)$ is *symmetric* if its value is unchanged by any permutation of its arguments. It is clear that the only continuous and symmetric functions f which satisfy the identity $\bar{f} \equiv 0$ are the n functions $O_r(x_1, \cdots, x_n)$. Hence we can state

THEOREM 2. *The only symmetric d. f. u. t. l.'s are the n functions* $O_r(x_1, \cdots, x_n)$ $(r = 1, \cdots, n)$.

A FORMULA FOR SAMPLE SIZES FOR POPULATION TOLERANCE LIMITS

By H. Scheffé and J. W. Tukey

Princeton University

In a paper to appear in a later issue of this journal dealing with various results on non-parametric estimation, we shall discuss in detail an approximate formula for the numerical calculation of sample sizes for Wilks' population tolerance limits. Because of the practical usefulness of this formula, it seems desirable to make it available without delay. Its accuracy is adequate for all direct applications.

An interval I is said to cover a proportion π of a univariate population with cumulative distribution function $F(x)$ if $\int_I dF = \pi$. Let $X_1, X_2, \cdots, X_n$ be a random sample from the population, and $Z_1 \leq Z_2 \leq \cdots \leq Z_n$ be a rearrangement of $X_1, X_2, \cdots, X_n$. Define $Z_0 = -\infty$, $Z_{n+1} = +\infty$, and consider the proportion B of the population covered by the random interval (Z_k, Z_{n-m+1}). Then $Pr\{B \geq b\}$ is independent of $F(x)$ if $F(x)$ is continuous[1], and equals $1 - I_b(n - r + 1, r)$, where $r = k + m$ and $I_x(p, q)$ is K. Pearson's notation for the incomplete Beta function.

Choose a confidence coefficient $1 - \alpha$, a pair of positive integers k, m, and a fraction b. The sample size n for which we can make the statement "the probability is $1 - \alpha$ that the random interval (Z_k, Z_{n-m+1}) cover a proportion b or more of the population" is then determined by the equation

$$I_b(n - r + 1, r) = \alpha, \tag{1.1}$$

where $r = k + m$. Our approximate solution is

$$n \doteq \tfrac{1}{4}\chi^2_\alpha(1 + b)/(1 - b) + \tfrac{1}{2}(r - 1), \tag{1.2}$$

where χ^2_α is the 100α percent point on the χ^2-distribution with $2r$ degrees of freedom. The required values of χ^2_α may be found for $\alpha = .1, .05, .025, .01, .005$ in Catherine Thompson's table [2]. For this range of α, and for $b \geq .9$, extensive numerical calculations indicate that the error of (1.2) is less than one tenth of one percent, and is always positive, that is, n is slightly overestimated by (1.2). We have not yet obtained an analytic proof of this statement, which refers to the difference from the exact (and, in general, non-integral) solution of (1.1).

As explained elsewhere [1], formula (1.2) may be used for Wald's solution of the multivariate case.

REFERENCES

[1] H. Scheffé, *Annals of Math. Stat.*, Vol. 14 (1943), p. 324.
[2] C. M. Thompson, *Biometrika*, Vol. 32 (1941), p. 189.

---

[1] That the theory is valid in this case we show later. Previous proofs have required the continuity of $F'(x)$.

A GENERALIZATION OF WARING'S FORMULA

By T. N. E. Greville

Bureau of the Census

Waring's formula (frequently, but less correctly, called Lagrange's formula) gives the polynomial of degree n taking on specified values for $n + 1$ distinct arguments. It is frequently used for interpolation purposes in dealing with functions for which numerical values are given at unequal intervals. This formula may be written in the form:

$$f(x) = \sum_{i=0}^{n} \left[f(a_i) \prod_{j(\neq i)=0}^{n} \frac{x - a_j}{a_i - a_j} \right], \tag{1}$$

where $a_0, a_1, a_2, \cdots, a_n$ are the arguments for which the value of the polynomial $f(x)$ is given. This formula was first published by Waring [2] in 1779, and it was not until 1795 that Lagrange gave it in his book: *Leçons Elèmentaires sur les Mathèmatiques*. The prominent British actuary and mathematician, Mr. D. C. Fraser states that "there are identities of notation in the statement of the formula which leave little doubt that Lagrange was simply quoting from Waring's paper." Waring's priority was brought to my attention by Mr. Fraser and by Dr. W. Edwards Deming.

If any two or more of the arguments a_i are equal, the form (1) becomes indeterminate. However, the limiting value, as $m + 1$ specified arguments approach a common value a, can be shown to be an expression involving the first m derivatives of the polynomial $f(x)$ for the argument a. This case of "repeated arguments" is of considerable interest, especially in connection with the theory of osculatory, or smooth-junction interpolation [1, p. 33]. It is the purpose of this note to generalize the formula (1) to the case in which not only the value of $f(x)$ but also of its first m_i derivatives are given for each argument a_i. The degree of the polynomial represented, which we shall denote by N, is $n + \sum_{i=0}^{n} m_i$. The generalized formula is:

$$f(x) = \sum_{i=0}^{n} \left[P_i(x - a_i) \prod_{j(\neq i)=0}^{n} \left(\frac{x - a_j}{a_i - a_j} \right)^{m_j+1} \right], \tag{2}$$

where $P_i(x - a_i)$ denotes a polynomial in $x - a_i$ obtained by the following procedure. First, $f(x)$ is expanded in a Taylor series in powers of $x - a_i$. Next, the expression $\left(1 + \frac{x - a_i}{a_i - a_j}\right)^{-m_j-1}$ is expanded as a binomial series for every j different from i. Finally, all the $n + 1$ expansions (n binomial and one Taylor) are multiplied together, and all terms containing powers of $x - a_i$ higher than m_i are rejected. This formula has already been given by Steffensen [1, p. 33] for the particular case in which every $m_i = 1$.

The general formula (2) is difficult to arrive at without a previous knowledge

of the result, but is easily shown to be the correct expression. Upon differentiating k times ($0 \leqq k \leqq m_r$) all the terms in the summation except the one corresponding to $i = r$ will contain the factor $(x - a_r)^{m_r - k + 1}$, and will therefore vanish for $x = a_r$. Moreover, the non-vanishing term, before differentiation, will agree, up to and including terms containing $(x - a_r)^{m_r}$, with the Taylor expansion of $f(x)$ in powers of $x - a_r$, since the product expression within the brackets will be exactly canceled, as far as terms of degree m_r, by the n binomial expansions. Hence the kth derivative of the non-vanishing term in the summation will be $f^{(k)}(a_r)$ for $x = a_r$. This establishes the formula.

This formula is clearly equivalent to the Newton divided difference interpolation formula with repeated arguments [1, p. 33], the argument a_i occurring $m_i + 1$ times. Therefore, if $f(x)$ is any function other than a polynomial of degree N or less, it is necessary to add a remainder term [1, pp. 22–23] of the form

$$f_N(x) \prod_{i=0}^{n} (x - a_i)^{m_i+1},$$

where $f_N(x)$ denotes the limiting value [1, pp. 20–21] of the divided difference of order N involving the arguments $x, a_0, a_1, \cdots, a_n$, with each argument a_i appearing $m_i + 1$ times. The existence of all the indicated derivatives is, of course, essential.

REFERENCES

[1] J. F. Steffensen, *Interpolation*, Baltimore, 1927.

[2] E. Waring, "Problems concerning interpolation," *Phil. Trans. Royal Soc.*, Vol. 69 (1779), pp. 59–67.

NOTE ON THE VARIANCE AND BEST ESTIMATES

By H. G. Landau

Washington, D. C.

The purpose of this note is to point out a certain relation between the variances, σ_1^2 and σ_2^2, of the random variables, x_1 and x_2, and the probabilities,

$$P_1(t) = Pr[\,|x_1 - E(x_1)| < t]$$

$$P_2(t) = Pr[\,|x_2 - E(x_2)| < t].$$

This is, if $\sigma_1^2 < \sigma_2^2$, then $P_1(t) > P_2(t)$ in at least one interval, $t_1 < t < t_2$.

A note by A. T. Craig [1] gave an example for which it was stated that $\sigma_1^2 < \sigma_2^2$ and $P_1(t) \leq P_2(t)$ for every t; but, as was pointed by Neyman [2], calculation of the probabilities involved shows the statement to be incorrect.

The present result provides a certain justification for the use of minimum variance estimates by assuring that no other estimate with the same mean can have, for every value of t, a greater probability of a deviation from the mean

less than t. If an estimate can be found which has a greater value of $P(t)$ for all t than does any other estimate, it is necessarily the minimum variance estimate.

The theorem below includes a similar relation for equal variances. This theorem can be obtained from known general results on inequalities for distributions determined by moments, [3] and [4]. The formulation given here with its significance for estimates does not appear to have been remarked.

THEOREM. *If the random variables, x_1 and x_2, have finite variances, σ_1^2 and σ_2^2, and*

$$\sigma_1^2 \leq \sigma_2^2,$$

then, either

$$Q(t) = P_1(t) - P_2(t),$$

is equal to zero at all points of continuity, which can occur only for $\sigma_1^2 = \sigma_2^2$, or there is an interval, $t_1 < t < t_2$, in which $Q(t)$ is positive.

PROOF. We write the variance as the Stieltjes integral,

$$\sigma_1^2 = \int_0^\infty t^2 \, dP_1(t),$$

and similarly for σ_2^2.

Let

$$S(T) = \int_0^T t^2 \, dP_1(t) - \int_0^T t^2 \, dP_2(t) = \int_0^T t^2 \, dQ(t)$$

$$= T^2 Q(T) - 2 \int_0^T tQ(t) \, dt,$$

integrating by parts.

Now

$$T^2[1 - P_1(T)] = T^2 \int_T^\infty dP_1(t) \leq \int_T^\infty t^2 \, dP_1(t),$$

and since σ_1^2 is finite, $\int_T^\infty t^2 \, dP_1(t) \to 0$ as $T \to \infty$, so that $\lim_{T \to \infty} T^2[1 - P_1(T)] = 0$, and similarly for $P_2(t)$.

Hence $T^2 Q(T) = T^2[1 - P_2(T)] - T^2[1 - P_1(T)] \to 0$ as $T \to \infty$, and since by definition $\lim_{T \to \infty} S(T) = \sigma_1^2 - \sigma_2^2$ it follows that

$$\sigma_1^2 - \sigma_2^2 = -2 \int_0^\infty tQ(t) \, dt.$$

From this it can be seen that either, $Q(t)$ vanishes at all points of continuity, in which case $\sigma_1^2 = \sigma_2^2$, or $Q(t)$ must be positive in some interval, since otherwise $\int_0^\infty tQ(t) \, dt$ must be negative and hence $\sigma_1^2 - \sigma_2^2 > 0$ contrary to the assumption, $\sigma_1^2 \leq \sigma_2^2$.

REFERENCES

[1] A. T. Craig, "A note on the best linear estimate," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 88–90.

[2] J. Neyman, *Math. Reviews*, Vol. 4 (1943), p. 280.

[3] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, McGraw-Hill (1937), pp. 373–380.

[4] A. Wald, "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments," *Trans. Amer. Math. Soc.*, Vol. 46 (1939), pp. 280–306.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of general interest

Personal Items

Dr. R. L. Anderson, on leave from the North Carolina State College, is serving as a research mathematician on a war research project at Princeton University.

Professor Kenneth J. Arnold, on leave from the University of New Hampshire, is doing war research in the Statistical Research Group, at Columbia University.

Dr. W. J. Dixon, on leave from the University of Oklahoma, is serving as a research mathematician on a war research project at Princeton University.

Mr. R. M. Foster of Bell Telephone Laboratories has been appointed professor and head of the department of mathematics of the Polytechnic Institute of Brooklyn.

Dr. Hilda Geiringer, Lecturer at Bryn Mawr College, has been appointed Professor of Mathematics and Head of the Mathematics Department at Wheaton College, Norton, Massachusetts.

Assistant Professor E. H. C. Hildebrandt of the State Teachers College, Upper Montclair, New Jersey, has been appointed to an assistant professorship at Northwestern University.

Mr. John F. Kenney of the University of Wisconsin has been promoted to an assistant professorship.

Assistant Professor L. A. Knowler of the University of Iowa has been promoted to an associate professorship.

Dr. Saul B. Sells is now Head Statistician of the Statistical Standards Branch, Office of Price Administration, Washington.

Mr. Walter H. Thompson, formerly an instructor at Virginia Polytechnic Institute, is now associated with the Kellex Corporation, New York.

Professor Helen M. Walker of Teachers College, Columbia University, has been elected president of the American Statistical Association.

Professor C. C. Wagner of Pennsylvania State College has been appointed assistant dean of the School of Liberal Arts.

Dr. Edward Helly of the Illinois Institute of Technology died November 28, 1943.

New Members

The following persons have been elected to membership in the Institute:

Andrews, T. Gaylord. Ph.D. (Nebraska) Instructor in Psychology, Barnard College, Columbia University, New York City.

Angell, Dorothy T. Member of Technical Staff, Bell Telephone Laboratories. *Murray Hill, New Jersey.*

Barnes, John L. Member of Technical Staff, Bell Telephone Laboratories. *21 North Cherry Lane, Rumson, New Jersey.*

Bloom, Rose. B.A. (Hunter) Pvt., Billings General Hospital, Ft. Benj. Harrison, Indiana.

Bonnar, Robert Underwood. M.S. (Univ. of Washington) Associate Chemist, Bureau of Ships, Navy Dept. *414 Whitestone Road, Silverspring, Maryland.*

Brearty, C. R. B.S. (California) Officer in Charge, Quality Control Section, Signal Corps Inspection Agency. *19 West 4th St., Dayton 2, Ohio.*

Campbell, George Clyde. M.S. (Iowa) Captain CWS Chief, Fiscal Div., Pine Bluff Arsenal. *Troy Road, RFD #1, Boonton, New Jersey.*

Clifford, Paul C. A.M. (Columbia) Asst. Prof. of Math., State Teachers College, Montclair, New Jersey. *541 Upper Mountain Ave.*

Cobb, William J. Statistician, Census Bureau. *4036 8th St., NE, Washington, D. C.*

Coggins, Paul Pond. M.A. (Harvard) Accountant, American Tel. and Tel. *195 Broadway, New York, N. Y.*

Dietzold, Robert L. Ph.B. (Yale) M.T.S., Bell Telephone Lab. *34 W. 11th St., New York, N. Y.*

Elconin, Victor. M.S. (California Inst. of Technology) Associate Physicist, California Inst. of Technology. *740 Cordova Ave., Glendale 6, California.*

Ferrell, Enoch B. M.A. (Oklahoma) Member of Technical Staff, Bell Telephone Laboratories. *75 Fuller Ave., Chatham, New Jersey.*

Ferris, Charles Duncan. A.B. (Princeton) Engineering Statistician, Surveillance Branch, T/3 Army. Ballistic Research Laboratory, Aberdeen Proving Ground, Md.

Goldberg, Henry. M.A. (Columbia) Asst. Mathematical Statistician, Columbia University, *401 West 118th Street, New York 27, N. Y.*

Gordon, Donald A. A.M. (Columbia) Assistant, Columbia University. *1327 E. 26 Street, Brooklyn, New York.*

Greenleaf, Herrick E. H. Ph.D. (Indiana) Prof. of Mathematics. *1024 S. College Avenue, Greencastle, Indiana.*

Griffitts, C. H. Ph.D. (Michigan) Professor of Psychology. *1507 Charlton Ave., Ann Arbor, Michigan.*

Hadley, Clausin D. Ph.D. (Wisconsin) Statistician, Marketing Research Dept., Eli Lilly and Company, Indianapolis 6, Indiana.

Halbert, K. W. A.M. (Harvard) Statistician, Amer. Tel. and Tel. Co., *195 Broadway, New York, N. Y.*

Hall, Marguerite F. Ph.D. (Michigan) Asst. Prof. Public Health Statistics. *25 Ridgeway, Ann Arbor, Michigan.*

Halmos, Paul Richard. Ph.D. (Illinois) Asst. Professor of Mathematics, Syracuse University. *513 Fellows Avenue, Syracuse 10, N. Y.*

Harold, Miriam S. B.A. (Hunter) Member of Technical Staff, Bell Telephone Laboratories. *19 Hillside Avenue, Chatham, New Jersey.*

Hatke, Sister M. Agnes. M.S. (Indiana State Teachers College) Graduate student at Purdue University. *St. Francis College, Lafayette, Indiana.*

Hizon, Manuel O. M.A. (Michigan) Philippine Govt. Scholar at Univ. of Michigan. *816 Packard, Ann Arbor, Michigan.*

Hodgkinson, William, Jr. B.A. (Harvard) Major, HQ AAF, U. S. Army. *195 Broadway, New York, N. Y.*

Jacob, Walter C. Ph.D. (Cornell) Lt., U. S. Navy. Bureau of Ships, Navy Dept., Washington, D. C.

Jones, Howard L. C.P.A. (Illinois) Supervisor of Revenue Results, Illinois Bell Telephone Co. *Room 1100, 309 W. Washington St., Chicago 6, Ill.*

Kaitz, Hyman B. A.B. (George Washington Univ.) Statistical Consultant, Psychological Research Unit No. 11, B.A.A.F., Fort Myers, Fla.

Keyfitz, Nathan. B.Sc. (McGill) Statistician, Dominion Bureau of Statistics, Ottawa, Canada. *Billings Bridge, Ontario.*

Kirchen, Calvin J. M.A. (Wisconsin) Product Control Statistician, U. S. Rubber Co. *4028 11th St. Place, Des Moines 13, Iowa.*

LaSala, Lucy Anne. B.A. (Hunter) Applied Math Group, Columbia Univ. *256 Irving Avenue, Brooklyn 27, New York.*

Leone, Fred Charles. M.S. (Georgetown Univ.) Instructor in Math., Purdue Univ. *310 N. Salisbury St., W. Lafayette, Indiana.*

McNamara, Kathryn J. M.A. (Clark Univ.) Economist, Business Research Dept., H. J. Heinz Company, P.O. Box 57, Pittsburgh 30, Pa.

McPherson, John Cloud. B.S. (Princeton) Director of Engineering, Int'l Business Machines Corp. *590 Madison Ave., New York 22, N. Y.*

Millikan, Max F. Ph.D. (Yale) Asst. Director, Div. of Ship Requits., War Shipping Admin. *2313 Huidekoper Pl., NW, Washington 7, D. C.*

Morrow, Dorothy Jeanne. M.S. (Univ. of Washington) Fellow in Mathematical Statistics, Columbia University. *605 W. 115 St., New York 25, N. Y.*

Morse, John W. M.A. (Columbia) Ordnance Engineer, Quality Control Unit Inspection Sec., Ammunition Branch, War Dept. *11 Verne St., Bethesda, Md.*

Mottley, Charles McCammon. Ph.D. (Toronto) Lieut., USNR, Bureau of Ships, (333), Navy Dept., Washington, D. C.

Murphy, Ray B. B.A. (Princeton) 2nd Lt., U.S. Marine Corps Reserve. *28 Godfrey Rd., Upper Montclar, N. J.*

Myslivec, Vaclav. Ph.D. (Prague) Czechoslovak delegate to the United Nations Interim Commission on Food and Agriculture. *Room 606, 1775 Broadway, New York 19, N. Y.*

Nicholson, George Edward, Jr. M.A. (Univ. of North Carolina) Asst. Mathematician, Applied Mathematics Group, Columbia Univ. *176 Park St., Montclair, N. J.*

Osterman, Herbert William. B.S. (Michigan) *c/o B. E. Wyatt, 1029 Vermond Ave. NW, Washington, D. C.*

Parke, Nathan Grier, III. A.B. (Princeton) Senior Aviation Design Research Engr., Navy Dept. *Malvern Ave., Ruxton 4, Md.*

Priestley, Alice E. A. M.A. (New York Univ.) Instructor in Math, Lafayette College. *226 McCartney St., Easton, Pa.*

Rapkin, Chester. M.A. (American University) Associate Statistical Analyst, Deputy Director, Division of Operating Statistics, Federal Home Loan Bank Admin., *2 Park Avenue, New York 16, N. Y.*

Rivoli, Bianca. M.A. (Columbia University) Statistician, Judson Health Center. *3265 Bainbridge Avenue, New York 67, N. Y.*

Roshal, Sol M. B.S. (Chicago) Officer in charge of research statistics, PRU #1, Nashville Army Air Center, Nashville, Tenn.

Ross, Frank A. Ph.D. (Columbia) Editor, Journal of the American Statistical Assn. *Thetford, Vermont.*

Schaeffer, Esther. A.B. (Chicago) Technical Asst., Univ. of Michigan. *547 Elm Street, Ann Arbor, Mich.*

Schilling, Walter. M.D. (Harvard) Asst. Clinical Professor of Medicine, Stanford University Hospital, San Francisco 15, California.

Shannon, Claude E. Ph.D. (M.I.T.) Member of Technical Staff, Bell Telephone Laboratories, *463 West Street, New York, N. Y.*

Sherman, Jack. Ph.D. (Calif. Inst. of Technology) Research Chemist, The Texas Co. *170 Church St., Poughkeepsie, New York.*

Smallwood, Hugh M. Ph.D. (Johns Hopkins) Asst. Dept. Head, Physical Research Dept., U. S. Rubber Co., Market & South Streets, Passaic, N. J.

Smith, Joan Thiede. B.S. (Minnesota) Accountant. *673 East Nebraska Ave., St. Paul, Minn.*

Smith, R. Tynes, III. Chief, Transport Economics Branch, Traffic Control Div., Office Chief of Transportation, A. S. F. *1001 16th St. S., Arlington, Va.*

Sobczyk, Andrew. Ph.D. (Princeton) Staff Member, Radiation Lab., Mass. Inst. of Technology. *32 Bow Street, Lexington 72, Mass.*

Thurstone, Louis Leon. Ph.D. (Chicago) Professor of Psychology, University of Chicago, Chicago, Illinois.

Toralballa, Leopoldo V. Ph.D. (Michigan) Special Instructor, University of Michigan. *1109 Willard St., Ann Arbor, Mich.*

Walsh, John E. B.S. (Notre Dame) Mathematician, Lockheed Aircraft Corp. *707 East Elk Ave., Glendale 5, Calif.*

Weber, Bruce Travis. M.A. (Columbia) Member of Technical Staff, Bell Telephone Laboratories, Marray Hill, New Jersey.

Weiner, Louis. A.M. (Harvard) Economist, U. S. Bureau of Labor Statistics. *4215 Russell Ave., Mt. Rainier, Maryland.*

Woodward, Patricia. Ph.D. (Pennsylvania) Associate Executive Secretary, Committee on Food Habits, National Research Council. *2101 Constitution Ave., Washington 25, D. C.*

REPORT ON THE WASHINGTON MEETING OF THE INSTITUTE

The second regional spring meeting of the Institute of Mathematical Statistics was held at George Washington University, Washington, D. C., Saturday and Sunday, May 6 and 7, 1944, jointly with a regional meeting of the American Statistical Association. The 383 registrants at the joint meeting included the following 76 members of the Institute:

Paul H. Anderson, R. L. Anderson, Theodore W. Anderson, Jr., Kenneth J. Arnold, Kenneth J. Arrow, 1st Lt., AC, Blair M. Bennett, Ernest S. Blanche, C. I. Bliss, Bonnar Brown, Joseph G. Bryan, Marjorie F. Buck, A. George Carlton, C. W. Churchman, William J. Cobb, Major A. C. Cohen, Jr., Edwin L. Crosby, Haskell B. Curry, J. H. Curtiss, B. B. Day, Scott Dayton, W. Edwards Deming, Philip Desind, W. J. Dixon, J. L. Doob, Will Feller, William C. Flaherty, Thomas N. E. Greville, E. J. Gumbel, Trygve Haavelmo, Margaret Jarman Hagood, Morris H. Hansen, Elvin A. Hoy, Leonid Hurwicz, Lt. (jg) W. C. Jacob, Alice S. Kaitz, Evelyn M. Kennedy, Lila F. Knudsen, H. S. Konijn, T. Koopmans, Anita R. Kury, Jacob E. Lieberman, Philip J. McCarthy, Francis McIntyre, William G. Madow, Lt. C. J. Maloney, Sophie Marcuse, John W. Mauchly, A. M. Mood, Vladimir A. Nekrassoff, Monroe L. Norden, H. W. Norton, Victor Perlo, A. C. Rosander, William Salkind, Marion M. Sandomire, Max Sasuly, Franklin E. Satterthwaite, S. B. Sells, L. W. Shaw, W. Arthur Shelton, Walter A. Shewhart, Harry Shulman, Blanche Skalak, John H. Smith, R. T. Smith, Arthur Stein, Joseph Steinberg, J. W. Tukey, Joseph L. Ullman, David F. Votaw, Jr., Capt. A. N. Watson, Frank M. Weida, Louis Weiner, S. S. Wilks, Patricia Woodward, Bertram Yood.

All sessions were held jointly with the American Statistical Association.

Professor Will Feller of Brown University acted as Chairman for the Saturday afternoon session. The following papers were presented:

1. *Elements of the Theory of Testing Hypotheses.*
   J. H. Curtiss, Jr., Navy Department
2. *Large Sample Tests of Statistical Hypotheses.*
   Abraham Wald, Columbia University

Professor Frank Weida of George Washington University acted as Chairman for the Sunday morning session. The following contributed papers were presented:

1. *On the Statistics of Sensitivity Data.*
   C. West Churchman and Benjamin Epstein, Frankford Arsenal and the University of Pennsylvania
2. *Simplified Plotting of Statistical Observations.*
   E. J. Gumbel, New School of Social Research
3. *Distribution of Sample Variances and Covariances of Normally Distributed Noncentral Variables.*
   M. A. Girshick, Department of Agriculture
4. *An Application of the Variate Difference Method to Multiple Regression.*
   Gerhard Tintner, Department of Agriculture
5. *Autocorrelation in London Temperature.*
   Horace W. Norton, Department of Commerce

Professor Samuel S. Wilks of Princeton University acted as Chairman for the Sunday afternoon session. The following papers were presented:

1. *Regression Problems in Time Series.*
   Tjalling Koopmans, Combined Shipping Adjustment Board
2. *Foundations of the Theory of Time Series.*
   J. L. Doob, Navy Department

The business meeting of the Washington Chapter of the Institute was held Sunday morning. The proposed Constitution of the Washington Chapter was ratified, and elections under this constitution were held for the first time. The following officers were elected:

William G. Madow, Census Bureau—3 year term, Secretary of the Program Committee 1944–45
Solomon Kullback, War Department—2 year term
Frank Weida, George Washington University—1 year term.

W. G. Madow
Secretary, Program Committee
Washington Chapter